

COMDEM[®] 2017



30th International Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM 2017)

Held in Preston and Grange-Over-Sands
10th - 13th July 2017

Edited by: I. Sherrington, A. Onsy, R. Rao, H. Brooks, J. Philip

Conference Proceedings:

COMDEM[®] 2017

**Condition Monitoring and Diagnostic
Engineering Management 2017**

Edited by:

I. Sherrington, A. Onsy, R. Rao, H. Brooks, J. Philip

Proceedings of the 30th International Conference on Condition
Monitoring and Diagnostic Engineering Management

Held in Preston and Grange-Over-Sands
10th - 13th July 2017

**30th International Conference on Condition Monitoring and Diagnostic Engineering Management
COMADEM 2017**

(10th – 13th July 2017)

Published by

Jost Institute for Tribotechnology
University of Central Lancashire
Preston PR1 2HE
Lancashire, UK.

The views expressed in the papers in these document are entirely those of individual authors. The organisers of COMADEM assume no liability for any injury or damage resulting as a consequence of the use of the information presented.

Copyright 2018: Jost Institute for Tribotechnology, University of Central Lancashire.
(Except where otherwise indicated.)

This work is protected by copyright. However, authors retain rights to use their article for teaching and for distribution within their organisation without the need to obtain further permission.

First printed edition: October 2018

ISBN: 9781909755154

FORWARD

This milestone 30th International conference on condition monitoring and machine diagnostics, in the form of COMADEM 2017, was held at two venues in the North West of England. The historic industrial city of Preston, which was at the heart of the industrial revolution in Britain in the 18th and 19th centuries and the quiet seaside town of Grange-Over-Sands in Cumbria on the Southern fringes of English Lake District, the with its picturesque views and Victorian resort atmosphere.

The Grange Hotel, built in 1866 in a neo-Italianate style, was used as the conference hotel. It provided an unusual, but convivial retreat away from everyday commitments that allowing delegates to consider and discuss a wide range of issues in machine surveillance that face engineers in today's rapidly developing world.

The conference programme itself covered a wide range of topics in condition monitoring and machine diagnostics and was supported by special sessions on "tribotronics" and "intelligent condition monitoring", as well as by presentations from commercial organisations working in condition monitoring and machine diagnostics . This technical programme was further complemented by evening social events providing informal networking opportunities outside the settings of the technical tracks.

We thank all the delegates attending COMADEM 2017 for their contributions and whether you attended COMADEM 2017 or not, we hope you find these proceedings a useful and convenient resource.

Prof I Sherrington

Dr A Onsy

Prof R Rao

COMADEM 2017 Conference Chairs

Conference Committee Members

Conference Chairs

Prof Ian Sherrington (Director of the Jost Institute for Tribotechnology)
Dr Ahmed Onsy (Senior Lecturer in Maintenance Engineering)
Prof Raj Rao (Director COMADEM International)

Organizing Committee (University of Central Lancashire)

Dr Hadley Brooks
Prof Edward Smith
Prof Bogdan Matuszewski
Dr Darren Ansell
Dr Gonzalo Garcia
Prof Lik Kwan Shark
Dr Wei Quan
Dr Martin Varley
Dr Yu Zhou
Dr Nathalie Renevier
Dr Sabuj Malik

Advisory Panel

| | |
|------------------|--------------------------------|
| Dr. David Baglee | University of Sunderland, UK |
| Dr. Fengshou Gu | University of Huddersfield, UK |
| Dr. Brian Shaw | Newcastle University, UK |
| Dr. Ling Wang | University of Southampton, UK |

Conference Management (University of Central Lancashire)

Elizabeth Roberts – UCLan Conference Officer
Ghazanfar Ali – Graduate Intern
Carl Berry – Graduate Student
Wilbert Sinzara – Graduate Student, School of Engineering
Bryan Cassels – Graduate Student, School of Engineering
Jephin Philip - Graduate Student, School of Engineering
Jesvin George – Graduate Student, School of Engineering
Ahmed N-S Abufroukh – Graduate Student, School of Engineering
Essa Anas – Graduate Student, School of Engineering

COMADEM 2017 Conference Partners:



Institution of Mechanical Engineers, UK



Center for Intelligent Maintenance Systems



The European Lubricants Industry Magazine



MFPT Society



COMADEM International



University of Maryland CALCE



PHM Prognostics Health Management Society



University of Huddersfield



Cranfield University, IVHM, Boeing



Aeronautical and Mechanical Testing Centre

CONTENTS

WELCOME TO COMADEM 2017

CONFERENCE COMMITTEE MEMBERS

SESSION PAPERS

1. Modelling, Analysis and Optimisation 1 (Session Chair Dr. Lilach Huber)

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| The Impact of Maintenance Duration on the Downtime of an Offshore Wind Farm - Alternating Renewal Process (<i>H. Seyr, A. Barros, M. Muskulus</i>) | 3 |
| Condition Monitoring of High-Sided Tractor-Trailer Units Under Gusty Crosswind Conditions (<i>A. Abdulwahab, R. Mishra</i>) | 8 |
| Contributions of Structural Health Monitoring to the Reliability of an Offshore Fixed Platform (<i>E. Etebu and M. Shafiee</i>) | 17 |
| Experimental and Simulation Study on Impact Faults and Vibration Characteristics of Gear Box (<i>H. Zhipeng, G. Tao, W. Chen, Z. Zhenqiao</i>) | 25 |
| Improved Multiwavelet Denoising using Local Sliding Window for Multi-Fault Detection of Rotating Machinery (<i>J. Yuan, J. Zhao, Y. Wei, Y. Zhou</i>) | 34 |

2. Sensor Technology and Damage Detection 1 (Session Chair Prof. Jing Lin)

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Monitoring of Water Contamination in Gearbox Lubricant Based on Vibration Analysis (<i>K. F. Brethee, F. Gu, A. Ball</i>) | 41 |
| Weak Fault Diagnosis of Bearings Using a Cyclic Deconvolution Method by Combining SMHD with MEDA (<i>Y. Miao, M. Zhao, J. Lin, J. Zhao, J. Wang</i>) | 52 |
| Quantitative Detection of Steel Surface Cracks Based on Eddy Current Pulsed Thermography (<i>J. Ma, X. Xu, D. Zhen, H. Zhang, Z. Shi</i>) | 61 |
| Cluster-Based Thresholding of Phased Array Ultrasound for Anomaly Detection in Weld Inspection (<i>B. Cassels, L. Shark, S. J. Mein, A. Nixon, R. Turner</i>) | 68 |
| A Novel Two-Stage Approach to On-Site Condition Monitoring (<i>V. Leavers</i>) | 76 |

3. Monitoring, Diagnosis, Prognosis and Health Management 1 (Session Chair Prof. Mike Barnes)

| | |
|----------------------------------------------------------------------------------------------------------------------------------|----|
| Wind Turbine Fault Detection by Monitoring its Performance Using High-Resolution SCADA Data (<i>E. Gonzalez, J. Melero</i>) | 83 |
|----------------------------------------------------------------------------------------------------------------------------------|----|

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Estimation Method of Static Contact Wire Height by Using Contact Force on Commercial Line (<i>T. Usuda and M. Ikeda</i>) | 92 |
| Feature Selection for Remaining Useful Life Prediction of Spur-Bearings (<i>K. de Calle, S. Ferreiro, E. Konde, I. Bravo, A. Arnaiz</i>) | 98 |
| Active Lubricant Condition Monitoring (<i>J. George, A. Onsy</i>) | 106 |
| Value Based Management in Centrifugal Compressors Maintenance (<i>X. Liang, F Duan, D. Mba, I Bennett</i>) | 118 |
| 4. Modelling, Analysis and Optimisation 2 (Session Chair Prof. Yimin Shao) | |
| Planetary Gearbox Diagnosis Based on Morlet Wavelet Filter and Time Synchronous Averaging (<i>X. Xu, M. Zhao, J. Lin, C. Yan, J. Zhou, J. Wang</i>) | 129 |
| Advanced Prognosis Method for Circuit Breaker Condition Monitoring (<i>N. Chen, R. Yu, J. Ruan, X. Zhang</i>) | 135 |
| Application of Stochastic Resonance and Variational Mode Decomposition in Rolling Bearing Fault Diagnosis (<i>J. Wang, S. Chen, C. Zhang, L. Deng, J. Hu</i>) | 144 |
| 5. Maintenance Engineering / Risk, Safety Assessment and Management (Session Chair Dr. Guanming Dong) | |
| How Can SMES Adopt a New Method to Advanced Maintenance Strategies? A Case Study Approach (<i>D. Baglee, U. Gorostegui, E. Jantunen, P.Sharma, J. Campos</i>) | 155 |
| The Role of Cultural Development When Improving Maintenance Practice in the Automotive Supply Chain (<i>D. Dixon, K. Robson, A. Wheatley, D. Baglee</i>) | 163 |
| The Role of ISO 55000 Standard in Asset Integrity (<i>I. Ihemegbulem, D. Baglee, A. Wheatley, S. Mattonen-Arola</i>) | 172 |
| 6. Monitoring, Diagnosis, Prognosis and Health Management 2 (Session Chair Asst. Prof. Hugo Andre) | |
| Incipient Fault Diagnosis and Monitoring System for Planetary Gear Train (<i>L. Gao, L. Wang, Y. Shao, L. Yin, M. Du, Y. Yang, C. Zheng</i>) | 179 |
| Spline Wavelet Based Filtering for Denoising Vibration Signals Generated by Rolling Element Bearings (<i>J. Kansanaho, K Saarinen, T. Kärkkäinen</i>) | 185 |
| Fault Diagnosis of Early-Stage Spalling Fault Based on a Hybrid Approach (<i>L. Wang, Y. Shao, M. Du, Y. Yang, F. Guo, Z. Cao</i>) | 192 |
| Bogie Speed Estimation and Signal Source Separation via Rail Vibration Analysis (<i>F. Gu, Y. Xu, F. Liu, A. D. Ball</i>) | 200 |

7. End User Applications & Maintenance in Industry 1 (Session Chair Prof. Andrew Ball)

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Wind Turbine Bearing Fault Detected with IAS Combined with Harmonic Product Spectrum (<i>A. Hugo, K. Ilyes, L. Quentin</i>) | 211 |
| A Comparison of Architectures and Evaluation Metrics for Streaming Machine Learning Algorithms in Industry 4.0 Applications (<i>C. Berry, G. Hall, B. Matuszewski, L. Shark</i>) | 220 |
| Measurement Uncertainty When Thermal Effects on a Machine Tool Using Non-Contact Displacement Transducers (<i>S. Gorton, S. Fletcher, D. Clough, A. P. Longstaff</i>) | 227 |
| Research on Spectrum Components of Planetary Gearbox Vibration Signals by Dynamic Modelling and Filtering (<i>L. Zhang, N. Hu, F. Gu, Z. Cheng</i>) | 237 |

8. Sensor Technology and Damage Detection 2 (Session Chair Dr. Mahmood Shafiee)

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Combining Model-Based Monitoring and a Physics of Failure Approach for Wind Turbine Failure Detection (<i>J. Tautz-Weinert, S. J. Watson</i>) | 247 |
| On the Feasibility of Inferring the Applied Mechanical Loading of a Conveyor System Test Rig From Monitored System Parameters (<i>O. Gebler, B. Hicks, A. Harrison, M. Barker</i>) | 256 |
| Assessing Uneven Milling Cutting Tool Wear Using Component Measurement (<i>Z. J. Ahmed, P. W. Prickett, R. I. Grosvenor</i>) | 270 |
| Potential Use Cases of LWIR Cameras for Automation and Process Monitoring (<i>B. Eichentopf, R. Baltes, N. Fietz, K. Nienhaus</i>) | 278 |

9. Monitoring, Diagnosis, Prognosis and Health Management 3 (Session Chair Dr. Hao Zhang)

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Condition Monitoring of Guiding Systems by Means of Model-Based Virtual Sensors (<i>M. Gonzalez, O. Salgado, J. Croes, B. Pluymers, W. Desmet</i>) | 287 |
| A Kernel Fisher Discriminant Framework Based on Kernel Entropy - Component Analysis for Fault Diagnosis of Rolling Bearings (<i>H. Zhou, T. Shi, W. Lai, G. Liao, J. Xuan</i>) | 296 |
| Drill Bit Wear Monitoring Based on Vibration Signal Analysis (<i>H. Rafezi, F. Hassani</i>) | 303 |
| Detection and Diagnosis of Reciprocating Compressor Faults Based on Modulation Signal Bispectrum Analysis of Vibrations (<i>U. Haba, K. Brethee, O. Hassin, F. Gu, A. Ball</i>) | 309 |

10. Modelling Analysis and Optimisation 3 (Session Chair Dr. Christoph Gerke)

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| A Novel Technique to Reduce Measurement Errors due to Flow –Sensor Interactions in Multi-Sensor Conductivity Probes (<i>D. Albarzenji, R. Mishra</i>) | 321 |
| Experimental Model-Based Approach to Integrated Prognostic and Health Management of a Non-Linear Liquid Level System (<i>A. Al-Khafaji, R. I. Grosvenor</i>) | 329 |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Effects of the Fluid Film on the Frequency Response Function of the Structure of Journal Bearings (<i>Y. Kang, H Zhang, D. Zhen, Z. Zhi, F Gu</i>) | 337 |
| Misalignment identification Based on Dynamic Time Warping for Planetary Gearbox (<i>Z. Shen, H. Li, D. Zhen, H. Zhang, Z. Shi, F Gu</i>) | 344 |
| 11. End User Applications & Maintenance in Industry 2 (Session Chair Dr. Jonas Albers) | |
| Effect of Manufacturing Method of a Centrifugal Fan Hub on its Heat Dissipation Characteristics (<i>T. Eshaafi, J. Swinton, T. Asim, J. Irons, R. Mishra</i>) | 355 |
| Abrasive Spur Gear Wear Prediction Model: Utilization of the Statistical Design of Experiment (<i>S. Raadnui, K Wangveera</i>) | 361 |
| Degradation-Based Preventive Maintenance Policy for Railway Transport Systems (<i>F. Dinmohammadi, B. Alkali, M. Shafiee</i>) | 368 |
| Condition Based Maintenance Decision Making: a Practical Approach for Marine Vessels (<i>L. G. Huber, S. Kunz, M. Dettling</i>) | 377 |
| 12. Sensor Technology and Damage Detection 3 (Session Chair Dr. Erkki Jantunen) | |
| Detecting Bearing Faults Using Ensemble Average Autocorrelation Based Stochastic Subspace Identification (<i>Y. Xu, P. A. Van Vuuren, X. Tang, F. Gu, A. Ball</i>) | 389 |
| Theoretical Elucidation of Pass Frequency for Multi-Flaws in a Roller Bearing and Precise Diagnosis Method Using Decision Tree and Support Vector Machine (<i>Y. Mizushima, L. Song, S. Ota, P. Chen</i>) | 396 |
| Monitoring Shaft Fatigue Failures Using Online Monitoring Techniques (<i>A. Abufroukh, Ahmed Onsy, Ian Sherrington</i>) | 403 |
| 13. Advanced Signal Processing and Big Data 1 (Session Chair Dr. Ming Zhao) | |
| Bearing Fault Feature Detecting Based on Nonlocal Means Denoising (<i>Y. Wang, S. He, S. Zhang</i>) | 419 |
| The Influence of B-Value Distribution in Damage Evaluation of Structural Material Based on AE Parametric Analysis (<i>M. T. I. Khan, S. Tsurumaru and R. Hidaka</i>) | 428 |
| A Self-Adaptive AE Signal Classification Method on Energy Domain (<i>Y. Zhou, D. Wang, L. Lin</i>) | 433 |
| 14. End User Applications & Maintenance in Industry 3 (Session Chair Dr. Lei Mao) | |
| Research of Optimal Placement of Sensor for Wind Turbine Based on Immune Algorithm (<i>Z. Song, H. Li, H. Zhang, D. Zhen, Z. Shi</i>) | 445 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings (<i>M. Hemmer, T. I. Wang, K. g. Robbersmyr</i>) | 450 |
| Design and Development of a Multipurpose Test Rig for Intelligent Lip Seals (<i>W. Sinzara, H. Brooks, I. Sherrington, A. Onsy, E. H, Smith M. J. Yanez, L. Marquez</i>) | 459 |
| An Intelligent Maintenance System for Driverless Cars (<i>Jephin Philip, Ahmed Onsy, Martin Varley</i>) | 467 |
| 15. Advanced Signal Processing and Big Data 2 (Session Chair Dr. Mohammed Khan) | |
| The Way Cyber Physical Systems Will Revolutionise Maintenance (<i>E. Jantunen, U. Gorostegui, U. Zurutuza, F. Larrinaga, M. Alnano, G. Di Orio, P. Malo, C. Hegedus</i>) | 479 |
| Nuisance Attribute Projection Based Channel Compensation Technique and its Application in Bearing Performance Degradation Assessment (<i>H. Jiang, J. Chen, G. Dong</i>) | 488 |
| Dynamic State Recognition Using CNN-RNN Processing Pipeline (<i>E. R. Anas, B. J. Matuszewski</i>) | 497 |
| Study on Feature Extraction of Bearing Incipient Weak Fault Based on Dictionary Learning (<i>G. Dong, J. Chen, H. Yuan</i>) | 505 |
| 16. Monitoring, Diagnosis, Prognosis and Health Management 4 (Session Chair Dr. Violet Leavers) | |
| The Impact of Effective Prognostic Techniques (Predicting Remaining Useful Life) on Successful Implementation of Total Productive Maintenance in the Power Industry (<i>A. Alseiari, P. Farrell</i>) | 517 |
| Fault Diagnosis of a Polymer Electrolyte Membrane Fuel Cell Using Bayesian Network (<i>L. Mao, L. Jackson, B. Davies</i>) | 526 |
| Weighted Narrowband Envelope Spectrum and its Application to Bearing Fault Diagnosis (<i>J. Duan, T. Shi, J. Xuan, H. Zhou</i>) | 534 |

Session 1

Modelling, Analysis and Optimisation 1

Session Chair Dr. Lilach Huber

The Impact of Maintenance Duration on the Downtime of an Offshore Wind Farm - Alternating Renewal Process (*H. Seyr, A. Barros, M. Muskulus*)

Condition Monitoring of High-Sided Tractor-Trailer Units under Gusty Crosswind Conditions (*A. Abdulwahab, R. Mishra*)

Contributions of Structural Health Monitoring to the Reliability of an Offshore Fixed Platform (*E. Etebu and M. Shafiee*)

Experimental and Simulation Study on Impact Faults and Vibration Characteristics of Gear Box (*H. Zhipeng, G. Tao, W. Chen, Z. Zhenqiao*)

Improved Multiwavelet Denoising using Local Sliding Window for Multi-fault Detection of Rotating Machinery (*J. Yuan, J. Zhao, Y. Wei, Y. Zhou*)

The Impact of Maintenance Duration on the Downtime of an Offshore Wind Farm - Alternating Renewal Process

H. Seyr¹, A. Barros², M. Muskulus¹

¹ Department of Civil and Environmental Engineering

² Department of Production and Quality Engineering, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway

ABSTRACT

Maintenance cost in offshore wind farms are one of the main reasons why it is still not competitive with its onshore counter part. To lower the price of energy, the authors want to improve the modeling of offshore wind farms in order to help improving the applied maintenance strategies. In this paper, we investigate the repair of wind turbine components as an alternating renewal process. The time it takes to repair (renew) a component in a turbine, directly influences the downtime of this turbine, or even larger parts of the wind farm. This downtime leads to production losses, which in turn raise the cost of energy. Previous investigations by two of the present authors showed that a variation in repair times has a significant influence of the production losses. In this paper, an alternate renewal process is investigated as tool to calculate the influence of the repair time duration on the production losses. We show that if we assume an exponential distribution of repair times, the distribution of the downtimes can be calculated analytically. For other distributions, the calculation has to be done numerically, after fitting the parameters of the distribution to the available data. Further work with other distributions and actual repair time distributions is planned.

Keywords: Offshore Wind Energy, Failure Modelling, Maintenance Optimisation, Stochastic Modelling, Alternating Renewal Process

Corresponding author: Helene Seyr (Helene.seyr@ntnu.no)

1. INTRODUCTION

The offshore wind industry is a growing industry in Europe and other continents. However, Operations and Maintenance (O&M) costs still account for a significant part of the total cost of energy [1]. Lowering these costs will lower the cost per kilowatt of power produced and hence make offshore wind sources competitive to other energy sources. In order to reduce the O&M costs, many approaches are currently followed in research. These include condition monitoring, optimizing the preventive maintenance, and inventing components that are more reliable.

In order to reduce the costs of corrective maintenance, repairs following a failure that could not be avoided by preventive measures, decision support tools are being developed by several research groups. These aim to help for wind farm (WF) operators and maintenance service providers in deciding which maintenance strategy to use in order to minimize costs or maximize availability. Most of these models rely on simulations of the operation phase of the WF. The operation of the WF, its availability and production as well as the maintenance costs, are influenced by several factors. These factors include the failure rates of wind turbine components, the distance of the WF from an onshore maintenance base, the weather at a given location, the types of maintenance vessels, number of vessels, the number of workers and the turbine model. Two of present authors started to investigate variations in the repair times and their influence on the unscheduled downtimes of the wind farm [2,3]. They have shown that a variation in the repair times can lead to a significant increase in downtimes and hence in production losses. These production losses in turn raise the cost of electricity produced.

In this paper, we want to present an approach based on a stochastic process named alternating renewal process to investigate this same influencing factor. Alternating renewal processes can be used, as soon as we assume random failure and repair times and a repair that is a renewal. The paper is structured as follows. Section 2 gives an overview of the methods used, the results of the analysis are presented in Section 3. Finally, we conclude the work and give an outlook to future work in Section 4.

2. METHODOLOGY

In order to define an alternate renewal process, we need to take some assumptions. First, we assume that the lifetime duration of a turbine component is random, and that it is following a given distribution. We further assume that the maintenance duration, the repair time, follows a different distribution. The distributions used in the paper are only examples and not based on actual wind park data. For the failures, we assume a constant failure rate, which leads to an exponential failure distribution. For many wind turbine components, the failure rate is assumed to follow a bathtub curve, a constant repair rate is then realistic for most of the lifetime [4-6]. Exponential distributions for the failures are common in the literature [7-12].

Also for the repair time, we use a constant rate for a first analysis. Since there has not been done any analysis on the actual distribution on repair times yet, this is a very strong assumption. However, this is a first analysis for a first example of a distribution. Other types of distributions should be investigated as well as fitting the distribution parameters to actual wind park data.

It is further assumed that each repair is a renewal, replacing the faulty component with an identical new one. A renewal is for many wind turbine components, like gears or electrical equipment a common form of repair.

The theoretical distribution for the downtime is calculated based on the distributions for the lifetime and repair times. We show that it is possible to do this calculation analytically for any failure and repair rate and the distribution of the downtime is presented dependent on the two rates.

3. RESULTS

In this section we present the results of our analysis. An example of an alternating renewal process is shown in figure 1. The turbine component that is considered has two different states – running or failed. In the given example, the component is running until time t_1 . At time t_1 , it fails and stays in state “failed” until its repair is completed at time t_2 . The process continues like this; the turbine is running from t_2 to t_3 , when it fails again. The next renewal is finished at t_4 , and so on. We are now interested in the distributions of the X_i and Y_i , the durations during which the component is running or failed respectively.

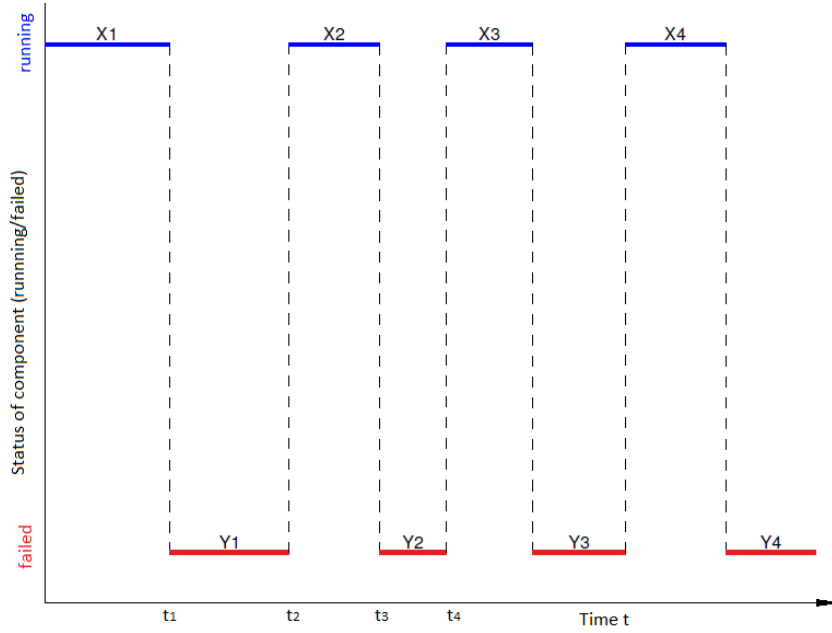


Figure 1. An example of an alternate renewal process. The X_i show the durations of operation (up states) and the Y_i show the downtimes (down states).

The failure dates (t_1, t_3, \dots) depend on the failure distributions and failure rate. As stated in the previous section, we are assuming an exponential distribution of the lifetime durations X_i , with a constant failure rate. Failure rates for different wind turbine components can be found in e.g. [5,6]. If we assume a known constant failure rate λ , the probability density function of the failure distribution can be written as:

$$f(t) = \lambda e^{-\lambda t} \quad (1)$$

We also assume an exponential distribution of the repair durations Y_i , which are corresponding to the downtimes, with a constant repair rate μ . The probability density function for this distribution is then:

$$g(t) = \mu e^{-\mu t} \quad (2)$$

The presented density functions represent the density of one failure and one renewal respectively. In order to know the distribution of a sum of lifetime durations $X_1 + \dots + X_n \sim ?$ we calculate the n -th fold convolution of $f(s)$ with itself:

$$F^{(n)}(t) = \int_0^t F^{(n-1)}(t-s) f(s) ds = \lambda e^{-\lambda t} \frac{t^{n-1}}{(n-1)!} \quad (3)$$

Since we are interested in the duration of the total downtime of the component, we introduce the following notation. $N(t)$ is the number of up states during $[0, t]$ and $D(t)$ is the amount of time spent in down state. We are now focusing on the calculation of the distribution of this $D(t)$, starting with the description of an event " $D(t) \leq x$ ". Since the total downtime is the sum of all downtimes, we have:

$$D(t) = Y_1 + \dots + Y_n \text{ if } N(t) = n. \quad (4)$$

The probability of an event " $D(t) \leq x$ " is then:

$$P(D(t) \leq x) = \sum_{n=1}^{\infty} P(Y_1 + \dots + Y_n \leq x | N(t-x) = n) \cdot P(N(t-x) = n). \quad (5)$$

Since when the failure occurs, n lifetimes have been completed, X_n has ended. As soon as X_{n+1} starts, the renewal has already taken place. Therefore the probability of having exactly n failures is the same as having a lifetime between $X_1 + \dots + X_n$ and $X_1 + \dots + X_{n+1}$. The probability $P(N(t) = n)$ can be simplified

$$P(N(t-x) = n) = P(X_1 + \dots + X_n \leq t-x \cap X_1 + \dots + X_{n+1} > t-x)$$

leading finally to

$$P(N(t-x) = n) = F^{(n)}(t-x) - F^{(n+1)}(t-x), \quad (8)$$

where $F^{(n)}(t-x)$ is as in Equation (3). For the conditional probability of having the downtime below a value x , while having exactly n failures we get

$$P(Y_1 + \dots + Y_n \leq x | N(t-x) = n) = G^{(n)}(x), \text{ where} \quad (9)$$

$$G^{(n)}(x) = \int_0^x G^{(n-1)}(x-s)g(s)ds, \quad (10)$$

and $g(s)$ the probability density function of Y_i . As for $F^{(n)}(t)$ in Equation (3), this simplifies to

$$G^{(n)}(x) = \mu e^{-\mu x} \frac{x^{n-1}}{(n-1)!} \quad (11)$$

Now the distribution of the downtime $D(t)$ can be calculated as

$$P(D(t) \leq x) = \sum_{n=1}^{\infty} G^{(n)}(x) (F^{(n)}(t-x) - F^{(n+1)}(t-x)) = e^{-\mu x - \lambda(t-x)} \sum_{n=1}^{\infty} \frac{\mu^n \lambda^n x^{n-1} (t-x)^{n-1}}{(n-1)!(n-1)!} \left(1 - \lambda \frac{t-x}{n}\right) \quad (12)$$

solving to

$$P(D(t) \leq x) = e^{-\mu x - \lambda(t-x)} \frac{\sqrt{\mu x(t-x)} I_0(2\sqrt{\lambda \mu x(t-x)}) - t\sqrt{\lambda} I_1(2\sqrt{\lambda \mu x(t-x)}) + x\sqrt{\lambda} I_1(2\sqrt{\lambda \mu x(t-x)})}{\sqrt{\mu x(t-x)}} \quad (13)$$

Where I_0 and I_1 are the modified Bessel functions [13] of the first kind of order 0 and 1 respectively.

4. CONCLUSIONS AND FURTHER WORK

In this paper, a first step has been taken to investigate the downtime of a wind turbine component as alternating renewal process. We have shown that for a constant repair rate, the distribution of the downtime can be calculated based on this repair rate and the corresponding failure rate of the component. This calculation is in principle possible for any repair duration distribution, but will need to be evaluated numerically in most cases, e.g. a Gamma distribution. The present authors are planning to investigate further distributions in the future and to use actual wind park data to conduct a case study.

ACKNOWLEDGEMENTS

The work of the first author has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska- Curie grant agreement No 642108. We would also like to thank the participants of the course PK8207 at NTNU in summer 2016 for their feedback on the very first ideas leading to this publication.

REFERENCES

- [1] Van Bussel G. and Schöntag C. (1997) Operation and maintenance aspects of large offshore windfarms EWEC- Conference, 272-275
- [2] Seyr H., Muskulus M. (2016) Interaction of repair time distributions with a weather model. Proceedings of the 29th International Congress on Condition Monitoring and Diagnostic Engineering Management, Xi'An, China.
- [3] Seyr H., Muskulus M. (2016) Value of information of repair times for offshore wind farm maintenance planning. Journal of Physics Conference Proceedings 753.
- [4] Tavner P., Xiang J, Spinato F. (2007) Reliability analysis for wind turbines. Wind Energy 10, 1–18.
- [5] Faulstich S., Hahn B. and Tavner P.J. (2011) Wind Energy 14, 327-337
- [6] Carroll, J., McDonald, A., Mcmillan, D. (2016) Failure rate, repair time and unscheduled O & M cost analysis of offshore wind turbines. Wind Energy 19, 1107–1119.
- [7] Endrerud O.E.V., Liyanage J.P. and Keseric N. (2014) Marine logistics decision support for operation and Maintenance of offshore wind parks with a multi method simulation model. Proceedings of the 2014 Winter Simulation Conference.
- [8] Hofmann M and Sperstad I B (2013) Energy Procedia 35, 177-186.
- [9] Scheu M., Matha D., Hofmann M. and Muskulus M. (2012) Energy Procedia 24, 281-288.
- [10] Douard F., Domecq C. and Lair W. (2012) Energy Procedia 24, 255-262
- [11] Dinwoodie I., McMillan D., Revie M., Lazakis I. and Dalgic Y. (2013) Energy Procedia 35, 157-166
- [12] Endrerud O.E.V. and Liyanage J.P. (2015) Decision Support for Operations and Maintenance of Offshore Wind Parks. Proceedings of the 8th World Congress on Engineering Asset Management (WCEAM 2013) & the 3rd International Conference on Utility Management & Safety (ICUMAS) (Cham: Springer International Publishing), 1125-1139
- [13] Arfken G.B., Weber H.J. and Harris F.E. (2013) Chapter 14 - Bessel Functions, Mathematical Methods for Physicists (Seventh Edition), Academic Press, Boston, 643-713

Condition Monitoring of High-Sided Tractor-Trailer Units Under Gusty Crosswind Conditions

Abubaker Abdulwahab*¹, Rakesh Mishra²

^{1,2} School of Computing & Engineering, University of Huddersfield, Queensgate,
Huddersfield, HD1 3DH, UK

ABSTRACT

It is well established that under gusty crosswind conditions, vehicle aerodynamic forces can lead to sudden changes in vehicle dynamics and stability. Large class vehicles, in particular, are more prone to rollover accidents in strong crosswind situations, especially at cruising speeds. It is therefore essential to conduct detailed investigations on the aerodynamic performance of commercial vehicles under crosswind conditions in order to improve their crosswind stability. This study predicts unsteady aerodynamic forces on a high-sided tractor-trailer unit exposed to gusty crosswind conditions. Although natural wind gusts are highly stochastic phenomena, and have a large variability in types and origins, this investigation suggests using deterministic gust model in combination with Computational Fluid Dynamics (CFD) based approach, using Large Eddy simulation (LES) for modeling air turbulence. Crosswind scenario with gusts of exponential shape, as specified in the Technical Specification for Interoperability (TSI), has been presented in the present study. A series of time-dependent aerodynamic interactions on the tractor-trailer unit have been recorded and investigated through the visualization of instantaneous gust flow structures around the truck-trailer unit. The results show that the TSI gust scenario has significant unsteady effects on the side aerodynamic force and the roll moment of the vehicle. Furthermore, there are significant variations in aerodynamic loads, consistent with the gust's strength.

Key words: Modelling, CFD, Deterministic gust

Corresponding author: Abubaker Abdulwahab (Abubaker.Abdulwahab@hud.ac.uk)

1. INTRODUCTION

The real atmospheric wind to which a ground vehicle is subjected during on-road driving is fully turbulent and unsteady. Within a turbulent wind field, strong wind pulses can occur and they are known as wind gusts or, more simply, gusts. Under crosswind gust conditions, sudden variations in wind loads have adverse impacts on the dynamic stability of road vehicles[1]. For instance, the aerodynamic roll moment works against vehicle weight loads, thus increasing the risk of rollover or derailment [2]. Therefore, for improving crosswind stability of ground vehicles under gusty wind actions, unsteady vehicle aerodynamic forces are among the essential information needed to carry out the investigation[3].

Modelling wind gust event is a complex task since in the very short-term, the gust speed estimation becomes infeasible. However, for design purposes, gusty wind condition can be evaluated based on the constrained simulation [4]. In this approach, extreme gusts are usually idealized as a deterministic gust, and then the gust is superposed to wind turbulent fluctuations[5]. Also, the deterministic gust model generally describes the stochastic character of the turbulence with the shape of occasional occurring wind peaks[6].

Gust models centered on deterministic approach are applied in various technical and scientific fields. In the rail vehicles area, a large European research projects TRANSAERO[7]and the DEUFRAKO crosswind program[8] contributed to standards in the form of the Technical Specification for Interoperability (TSI) [9,10]. According to these standards, the wind scenario prescribes a deterministic wind gust based on a bi –

exponential function, also known as “Chinese hat”. A number of studies have utilized the TSI gust scenario to produce the extreme wind loads on rail vehicles (e.g. [5], [11] and [12]).

Once the wind speed time-history is produced as deterministic gust scenario, extreme wind loads on a vehicle are usually expressed through aerodynamic coefficients as [13]:

$$F(t) = \frac{1}{2} \rho_{\text{air}} A_r C_i V_{\text{rel}}^2(t), i = D, S, L \quad (1)$$

$$M_{x,y,z}(t) = \frac{1}{2} \rho_{\text{air}} A_r h C_{mi} V_{\text{rel}}^2(t), i = D, S, L \quad (2),$$

where F is a generalized aerodynamic force and M is a generalized aerodynamics moment, ρ_{air} is the air density, A_r is a reference area and h is a reference height. The aerodynamic coefficients are for drag force C_D , lift force C_L , and side force C_s . In Eq. (2) the coefficients are rolling moment coefficient (C_{mD}), pitching coefficient (C_{mS}) and yawing coefficient (C_{mL}). The wind speed relative to the vehicle (V_{rel}) can be defined directly as a function of the absolute wind speed time-history $u(t)$ seen by the vehicle with velocity of V_{tr} , and it is defined as [3]:

$$V_{\text{rel}}^2(t) = (V_{tr} + u(t) \cos \beta)^2 + (u(t) \sin \beta)^2 \quad (3),$$

where β is a wind angle relative to the vehicle moving direction.

Generally, three techniques are conducted for estimating steady and unsteady aerodynamic coefficients of ground vehicles: full-scale measurements, scale model experiments and computational fluid dynamic (CFD) simulations. It is extremely costly and difficult to set up a wind tunnel test that reproduces the ground relative motion [14]. In contrast, CFD computation is well incorporated into the automotive sector and is currently an engineering tool use correspondingly with tests during the aerodynamic design process of ground vehicles [15]. Moreover, the CFD technique can give great quantities of transient data and comprehensive three-dimensional information concerning the wind flow domain [16]. This information can help to elucidate the comprehensive mechanism of the unsteady aerodynamics of road vehicles.

Based on CFD simulation, several techniques were utilized in previous CFD investigations concentrating on turbulent models such as $K-\epsilon$, $K-\omega$, DES, and LES. However, few examples of using a deterministic gust model in combination with CFD approach can be found in the literature. Favre [15] carried out CFD study to analyses the flow around a simple car geometry under crosswind gust condition. In his investigation, a deterministic wind gust represented by a continuous and smooth step-like function has been used. Vehicle aerodynamic forces were obtained using a commercial software, and the results have been compared against two different types of mesh. His work reveals that deterministic gust models can be applied to simulate crosswind aerodynamic forces on a ground vehicle based on CFD method. However, the study was conducted on a simple small car model, which is less sensitive to the crosswind comparing with trains or high-sided lorries [1]. In addition, the author recommends further investigation based on other types of wind gusts. Similar to [15], a rectangular crosswind profile was considered by Tsubokura et al. [16]. The study carries out a numerical simulation of unsteady aerodynamics of full-scale truck model under wind gust by implementing the LES technique. Numerical values of drag and side coefficients have been validated by wind tunnel tests, and the results show good agreement. However, the maximum gust strength that has been applied was 4.43 m/s, which is relatively small comparing to the strength of the gust in reality.

The TSI wind scenario that has been applied to rail vehicles represents a train traveling on an embankment under constant mean wind load and suddenly being hit by an extreme wind gust [12]. This situation applies to road vehicles as well, and from a practical point of view, it is of interest to examine the TSI gust in the road vehicle aerodynamics. Therefore, the present paper is an attempt to use the CFD simulations in combination with the TSI deterministic gust scenarios. The possibility to apply the gust on high-sided trailer is investigated based on the LES turbulent technique. The gust scenario is introduced into the CFD

simulation by creating an external velocity data file according to the TSI gust equation and imposing this to the lateral inlet of the computational domain. The tractor-semitrailer model is fixed on the numerical domain floor, and uniform main flow is imposed at the main inlet, which represents the relative velocity of the surrounding air on the vehicle moving forward in a straight line.

2. METHOD

This section includes a summary of the CFD method that is implemented to analyse the vehicle aerodynamic forces. The CFD simulation was produced using ANSYS-Fluent 17.0 software.

2.1. Vehicle Model and Computational domain

A full-scale model of a tractor semi-trailer based on geometry of real commercial vehicle was developed, and it was adapted by eliminating each small part attached to its surface. As shown in figure 1, its length, width, and height are 19.03 m, 2.4m, and 4.16m, respectively. The vehicle is anticipated to move directly ahead at a continuous speed of 30 m/s, which is the maximum speed limit for commercial vehicles on highway according to several legislations.

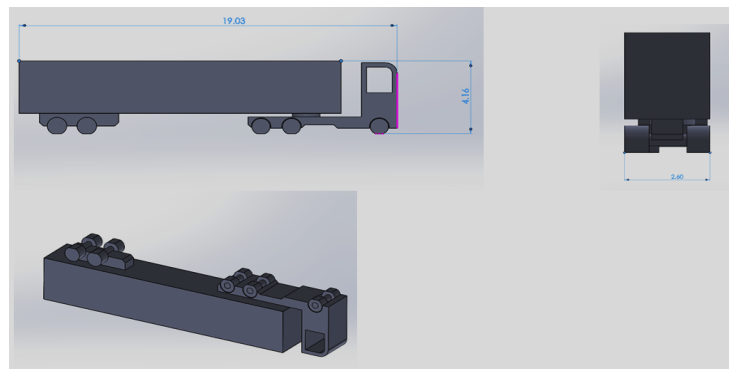


Figure 1. Vehicle 3D CAD model

Further, figure 2 illustrates a framework of the computational domain used in all CFD analysis presented in this paper. It is expressed as a rectangular duct similar to analysis domain that used in [16], with a 124.5 m length, 126.0 m width, and 32.0 m height. This big domain was implemented to capture the crucial flow features. The domain was split to about 16,250,000 elements with about 2,970,000 vertexes, the number of elements has been chosen after several meshes were tested to check the grid independency. The final mesh was tetrahedral unstructured mesh, this type of mesh has been used in the past for the LES in [16], and the DES in [15], with success. The topology of these control volumes is illustrated in figure 3. In this figure, two refinement zones are shown, the finest cells zone and the upstream zone. The fine cells are created to capture the small flow structures around the vehicle. Further, a constant time step of $\Delta t = 14 \times 10^{-4}$ s was used to achieve a Courant-Friederich-Lewy (CFL) number beneath 1 so as to acquire a correct and steady solution in this simulation. 20 iterations were performed for each time step to reduce the magnitude of non-linear residuals.

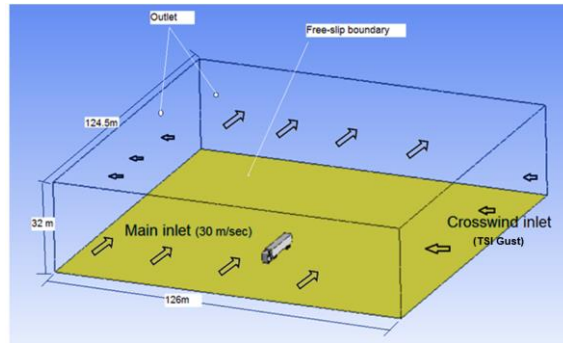


Figure 2. Computational domain and boundary conditions

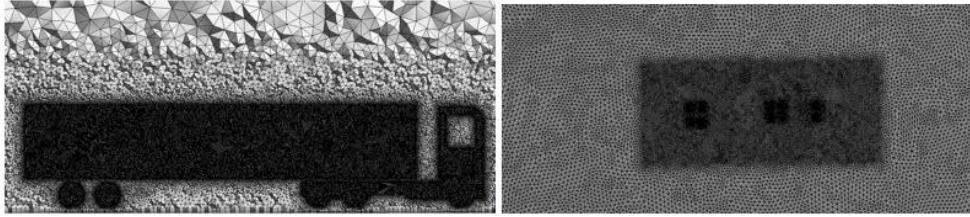


Figure 3. Numerical grids around the truck

2.2. Boundary conditions and boundary data

Figure 2 shows the boundary conditions that used in this investigation. Under these conditions, the truck model is affixed to the floor roughly 40 meters downstream of the domain inlet boundary. The ground is a moving wall with 30 m/s in the streamwise direction, which represents the vehicle speed, and the same speed is applied to the uniform flow at the front inlet. At the outlet faces, zero pressure is arranged, and a symmetry boundary condition is set at the top boundary. As previously mentioned, to reproduce a strong crosswind across the pathway of the truck, the TSI gust profile was imposed on the side boundary of the domain.

2.2.1 TSI Gust-characteristic

The TSI wind gust scenario is developed based on the analysis of real wind data [17]. The full TSI gust scenario (figure 4) involves a linear rise (from t_1 to t_2) to the base level of the wind speed \bar{u} when the truck semi-trailer is traveling at a consistent state (from t_2 to t_3). The increase of wind velocity aligns with the exponential gust profile that signifies the wind gust from t_3 to t_4 . Between t_4 and t_5 the wind velocity reduces to the preceding base level following the gust function. Finally, an additional time stage involving a constant wind speed takes place between t_5 and t_6 .

In a stationary wind speed, the gust is a perturbation velocity, u' , added on top of the mean wind speed, \bar{u} , depending on the Reynolds decomposition:

$$u(t) = \bar{U} + u'(t) \quad (4)$$

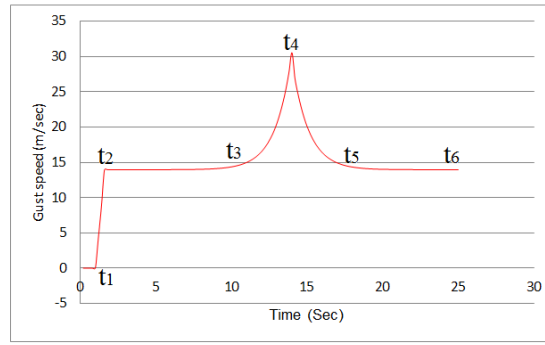


Figure 4. TSI gust scenario

According to the gust modeling methodology in the TSI standard, the perturbation velocity is modeled based on how the mean gust can be described by the autocorrelation function $C(t)$ and amplitude A , i.e $u' = A C(t)$. This leads to a relatively sharp waveform shape; the shape that has been shown to agree with the mean gust obtained from measurements [26]. Thus, wind speed of the bi-exponential profile shown in the figure 4 is calculated by the following formula [9].

$$u(t) = \bar{u} + \tilde{A} \cdot \sigma_u \cdot C(t) \quad (5),$$

where the average wind speed and the turbulence intensity determines the standard deviation of longitudinal element (following the average wind speed) of the wind:

$$\sigma_u = I \cdot \bar{u} \quad (6)$$

The value of normalized gust amplitude (\tilde{A}) is fixed by the technique as 2.84. The complex factor in Eq. 5 is the autocorrelation function ($C(t)$) which represents the correlation between the gust amplitude at the present moment and at the maximum amplitude. This function is formulated by

$$C(t) = \exp(-\sqrt{(C_u^x P_u^x)^2 + (C_u^y P_u^y)^2}) \quad (7),$$

where C_u^x and C_u^y are the coherence decay coefficients in the mean wind direction and perpendicular to the mean wind direction respectively; P_u^x and P_u^y are the exponential coefficients in the mean wind direction and perpendicular to the mean wind direction respectively. The values of these coefficients measured experimentally, and they are given in the TSI standard. After substituting the coefficients values in Eq. 7, the correlation function can be calculated as:

$$C(t) = \exp(-\sqrt{(5u_x)^2 + (16u_y)^2}) \quad (8)$$

As the wind gust is fixed in space, the transformation to calculate the temporal distribution is only possible when the vehicle speed is constant [9]. Then, velocity components of the wind along and cross the road are [10]:

$$u_x(\tilde{x}) = \frac{1}{2} \tilde{x} \cdot \cos(\beta) \cdot \frac{1}{T \cdot \bar{u}} \quad (9)$$

$$u_y(\tilde{x}) = \frac{1}{2} \tilde{x} \cdot \sin(\beta) \cdot \frac{1}{T \cdot \bar{u}} \quad (10),$$

where the \tilde{x} term in the equations is a function of the distance along the road towards the position of the maximum amplitude of the gust, and it can be calculated by [9]:

$$\tilde{x} = v_{tr}(t - t_{max}) \quad (11)$$

Further, by substituting u_x and u_y in Eqs. 8, for $\beta=90$ the coherence function is simplified to [10]:

$$C_{(\beta=90)} = e^{16u_x} \quad (12)$$

2.3. Governing equations

In principle, the aerodynamic forces are a result of interaction between a vehicle's body and the air surrounding it. To date, LES has been implemented successfully to assess aerodynamics performance of road vehicles, as may be evidenced in the literature (e.g.[3],[16],[14]). In this paper, an incompressible Newtonian fluid was supposed and the equations of continuity and momentum were spatially filtered to acquire the governing equations of LES. For further information on the LES turbulent model see reference[16].

3. SIMULATION RESULTS

For solving the pressure–velocity coupling, the results of the present simulations are obtained with the SIMPLE method. The spatial discretization schemes are a second order for the pressure equation and a bounded central difference for the momentum equation. Also, the bounded second-order implicit scheme is chosen for the transient terms.

3.1. Flow field visualization

In order to discuss trailer crosswind aerodynamic forces and moments, flow field around the trailer predicted in LES analysis is visualized. Figure 5 shows snapshot of gusty wind distribution on a horizontal plane at the mid-height of the trailer, and figure 6 depicts vortex structures around the truck body. The figures also show values of maximum gust velocity, which corresponding to the maximum gust velocity at 14.6 sec. At the maximum gust speed, the flow accelerates dramatically over the trailer's top and bottom surfaces and through the gap between the tractor and trailer. In addition, the flow separates to form large unsteady structures and fluctuate on both sides of the trailer.

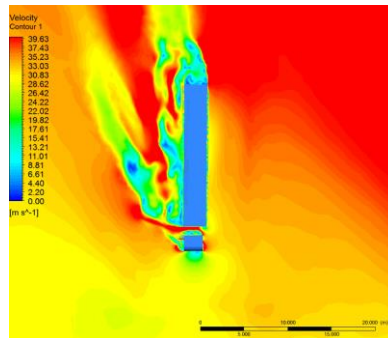


Figure 5. Contour of gust velocity magnitude on a horizontal cut-plane under the TSI gust at t=14.6 sec.

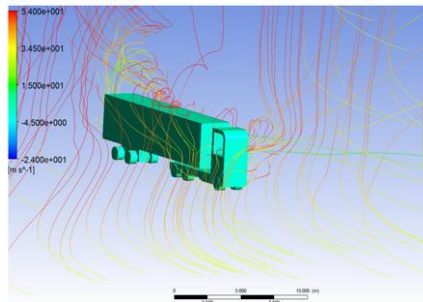


Figure 6. Instantaneous flow structures of the TSI gust around the sides of the trailer at t=14.6 sec.

Contours of the total pressure distributions (figure 7) have been computed for the leeward and windward sides of the vehicle. Due to the development of flow vortices on the leeward side, a low-pressure region is observed. The existence of the lower pressure region on the leeward side of the train explains the increase in the aerodynamic side force and roll moment. Moreover, high-pressure regions spreads widely on the windward side of the trailer at high speed.

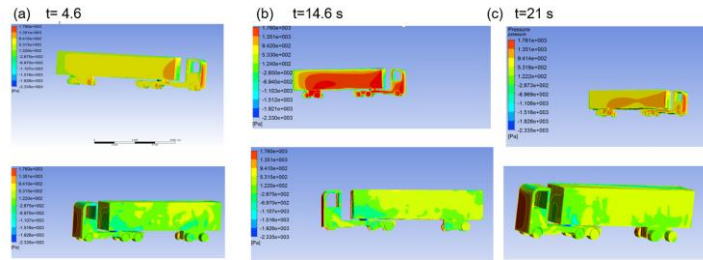


Figure 7. Contours of the total pressure distributions on the truck surfaces under TSI gust scenario (Top windward sides, bottom leeward sides)

3.2. Unsteady aerodynamics loads

Eqs. (1), (2) and (3) were used to calculate the unsteady aerodynamic forces acting on the trailer. Figure 8 illustrates these forces acting on the truck under TSI gust effects. As can be seen from this figure, the aerodynamics force monotonically increase and decrease consistent with the scale of the gust speed. The gust effects observed on the aerodynamic side force is practically about 2.5 times its effects on the drag and lift forces. Also, the vehicle reaches the exponential part of the gust scenario at approximately $t=8$ seconds, and subsequently, the side aerodynamic force rises from about 11 kN to maximum 33 kN in approximately 6 seconds.

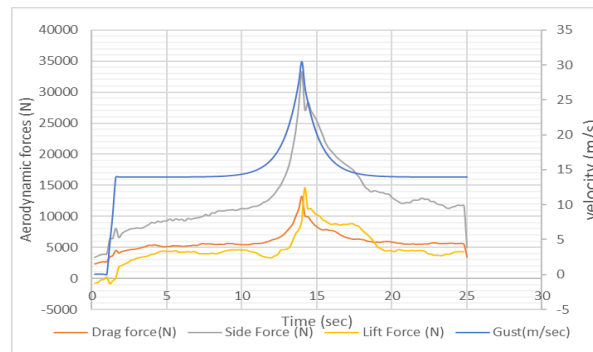


Figure 8. Time history of truck unsteady aerodynamic forces under the effect of the TSI wind scenario.

The gust wind effects can be more visibly perceived in the reaction of aerodynamic moments, figure (9). Amongst these, the rolling moment displays severe fluctuations, and the yaw moment also displays the penchant of instability. These sudden changes in aerodynamic moments are particularly notable and valuable for the appraisal of vehicular movement stability.

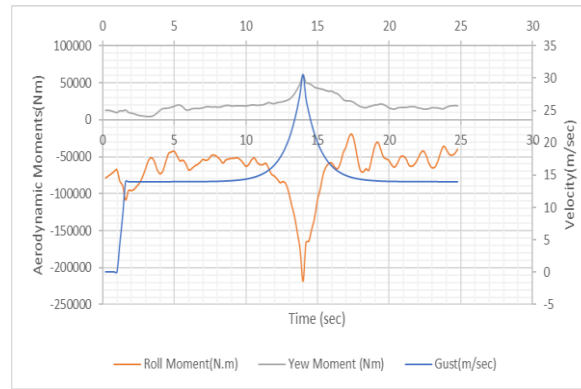


Figure .9 Duration of the unsteady aerodynamic rolling and yawing moments in the TSI wind scenario

4. CONCLUSION

The TSI gust scenario was applied to the full-scale tractor-semitrailer model in combination with CFD approach. The consequences of unsteady aerodynamics when the vehicle is exposed to this extreme gusty crosswind were predicted. The transient aerodynamic side force and rolling moment were observed to be significantly higher than other aerodynamic forces. Further, throughout the rushing in and out process, variability consistent with the scale of crosswind speed was discovered. These conclusions strongly suggested the significance of considering the unsteady aerodynamic forces in the analysis of heavy vehicle roll dynamics. Moreover, the simulations used in this study were limited to constant gust amplitude. The allowance of varying amplitudes, however, does not result in further problems. Moreover, it appears to be feasible to extend this technique to both bicycles and motorbikes traveling in wind-exposed environments.

REFERENCES

- [1] C. Baker, "Ground vehicles in high cross winds part III: The interaction of aerodynamic forces and the vehicle system," *Journal of fluids and structures*, vol. 5, no. 2, pp. 221-241, 1991.
- [2] F. Cheli, F. Ripamonti, D. Rocchi, and G. Tomasini, "Aerodynamic behaviour investigation of the new EMUV250 train to cross wind," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 98, no. 4, pp. 189-201, 2010.
- [3] C. Baker, F. Cheli, A. Orellano, N. Paradot, C. Proppe, and D. Rocchi, "Cross-wind effects on road and rail vehicles," *Vehicle system dynamics*, vol. 47, no. 8, pp. 983-1022, 2009.
- [4] P. Cheng, and W. Bierbooms, "Extreme gust loading for wind turbines during operation," *Journal of Solar Energy Engineering*, vol. 123, no. 4, pp. 356-363, 2001.
- [5] C. Proppe, and C. Wetzel, "A probabilistic approach for assessing the crosswind stability of ground vehicles," *Vehicle System Dynamics*, vol. 48, no. S1, pp. 411-428, 2010.
- [6] C. Knigge, and S. Raasch, "Improvement and development of one-and two-dimensional discrete gust models using a large-eddy simulation model," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 153, pp. 46-59, 2016.
- [7] B. Schulte-Werning, A. Deutsche Bahn, F.-u. Technologie-Zentrum, and B. Schulte, "The TRANSAERO Project-Joint European Railway Research on Transient Aerodynamics," *TRANSAERO: A European Initiative on Transient Aerodynamics for Railway System Optimisation*, vol. 79, pp. 11, 2013.
- [8] D. Consortium, "Common DEUFRAKO research on cross wind effects on high speed railway operation 2001-2004," *Final Report of DEUFRAKO SIDE WIND project*, Draft version, vol. 1, 2004.
- [9] O. J. o. t. E. Union, "Technical specification for interoperability," *Effects of Crosswinds*, 2008.
- [10] B. E. 14067-6:2010, "BS EN 14067-6:2010: Railway applications. , " *Aerodynamics. Requirements and test procedures for cross wind assessment*, Generic, British Standards Institute, 2010.
- [11] X. Zhang, *Crosswind stability of vehicles under nonstationary wind excitation*: KIT Scientific Publishing, 2015.
- [12] Wetzel, C., & Proppe, C. (2008, July). Crosswind stability of high-speed trains: a stochastic approach. In *BBA VI International Colloquium on Bluff Bodies Aerodynamics and Applications, Milano, Italy*.
- [13] A. J. Scibor-Rylski, *Road vehicle aerodynamics*, 1984.
- [14] R. McCallen, F. Browand, and J. Ross, *The aerodynamics of heavy vehicles: trucks, buses, and trains*: Springer Science & Business Media, 2013.
- [15] T. Favre, and G. Efraimsson, "An assessment of detached-eddy simulations of unsteady crosswind aerodynamics of road vehicles," *Flow, Turbulence and Combustion*, vol. 87, no. 1, pp. 133-163, 2011.

- [16] Tsubokura, K. Takahashi, T. Matsuuki, T. Nakashima, T. Ikenaga, and K. Kitoh, *HPC-LES for Unsteady Aerodynamics of a Heavy Duty Truck in Wind Gust-1st Report: Validation and Unsteady Flow Structures*, 0148-7191, SAE Technical Paper, 2010.
- [17] A. Carrarini, "Reliability based analysis of the crosswind stability of railway vehicles," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 95, no. 7, pp. 493-509, 2007.

Contributions of Structural Health Monitoring to the Reliability of an Offshore Fixed Platform

E. Etebu and M. Shafiee
Cranfield University, Bedfordshire MK43 0AL, UK

ABSTRACT

In this paper, the contribution of structural health monitoring (SHM) to the reliability of an offshore fixed platform is evaluated. A numerical analysis is carried out to determine the updated fatigue reliability index via fracture mechanics, for a structure member with three varying degrees of damage associated with crack size using two SHM systems with varying sensor density. It is observed that the updated reliability index of a structure member in a fixed offshore platform varies based on the SHM system's capability in detection of damages. The SHM system with lower probability of detection (POD) generates an updated reliability index larger than the initial reliability index throughout the structures operational life, in the presence of damage that is below its detectable size. However when the level of damage is greater than its detectable size, the updated reliability index plunges below the initial reliability index in the later portions of the structure's operational life. The second SHM system with a higher probability of detection, detects the three damage levels applied in this work hence, its updated reliability index indicated an overestimation of the structure member reliability index at an earlier time.

Keywords: Structural Health Monitoring, Sensors technology and damage detection, Monitoring, diagnosis, prognosis and health management.

Corresponding author: Mahmood Shafiee Other (email: m.shafiee@cranfield.ac.uk)

1. INTRODUCTION

Fixed jacket platform structures used in the offshore exploration of oil and gas are subjected to operational and environmental loads such as deck load, ship impact, waves, current, wind, and seismic activity, which lead to the buildup of damage on the structure during its operational life. In order to prevent catastrophic failure, the ability to detect and locate damage on a structure is vital. Due to the cyclical nature of environmental offshore loads applied over the structures operational life, it is of essence to assess a structures resistance to fatigue damage. Fatigue damage is generally evaluated using a fatigue limit state (FLS), this provides a relationship between the effect of operational and environmental loads to the structures fatigue resistance (Vughts & Kinra 1976). Uncertainties from physical, statistical, and model parameters arises in the formulation of FLS, thus reliability analysis is performed to determine fatigue reliability index, which indicates the probability that fatigue failure will not occur. The ability to accurately assess fatigue reliability is paramount to understanding the state of a structure, and determination of an inspection schedule. The effect of inspection on fatigue reliability index is generally conducted by updating the failure reliability index based on the outcome of a scheduled inspection (Madsen et al. 1987, Moan & Song 2000). Structural health monitoring (SHM) is a tool used to detect, localize, and analyze damage present in a structure by continuously monitoring the structures behavior, thus providing real time information of the structures integrity rather than periodic structural integrity attained from scheduled inspection (Doebeling et al. 1996). Previous works on SHM for offshore structures have focused on improving the damage detection, localization and assessment; however the contribution of SHM to offshore structural reliability has not been addressed. In this work, we assess the contribution of SHM to the fatigue reliability of an offshore fixed jacket platform structure member, using an updated fatigue reliability index generated from the outcome of SHM. The initial fatigue reliability index was calculated using Advanced First Order Second Moment method (AFOSM), where FLS was defined using fracture mechanics. Updated reliability index is calculated using a conditional probability of failure based on FLS and SHM event detection. SHM damage detection and localization was conducted using modal frequencies generated from

a 3D fixed jacket platform model, with varying levels of damage. This current work is limited to theoretical and numerical analysis, applications to experimental and real world offshore jacket platforms represents the future direction of this work.

2. FATIGUE LIMIT STATE FUNCTION

Fatigue limit state function was described based on fracture mechanics where Paris-Ergodan equation relates the rate of crack growth per load cycle and range of stress intensity factor ΔK as

$$\frac{da}{dN} = C[\Delta K]^m, \quad \Delta K > 0 \quad (1)$$

where parameters C and m are material properties. The range of stress intensity factor is generally expressed as a function of crack size a , range of stress ΔS , and structural geometric function Y using

$$\Delta K = Y\Delta\sigma\sqrt{\pi a} \quad (2)$$

A combination of equation 1 and 2, arises in the change of crack size over a structures service life for a total number of stress cycles N as expressed in equation 3, where a_o is the initial crack depth, a_n is the crack depth after N stress cycles, and $E[\Delta S^m]$ is the stress range distribution.

$$\int_{a_o}^{a_N} \frac{da}{Y^m[\pi a]^{0.5-m}} = C \cdot N \cdot E[\Delta S^m] \quad (3)$$

To describe the vast number of long term distribution of the range of stress on an offshore structure at varying sea states, Weibull distribution fitted to multiple short term stress ranges which results in equation 4 (Pillai & Prasad 2000, Karadeniz 2001, Shabakhty 2004). Parameter A represents scale functions of the Weibull distribution, Γ is a gamma function, and B is the Weibull distribution shape function.

$$E[\Delta S^m] = A^m \cdot \Gamma\left[1 + \frac{m}{B}\right] \quad (4)$$

In order to represent the change of crack size over a structures service life with time, the total number of stress cycles over a structures service life can be expressed as a function of the time history of stress T , and the average frequency for long term stress range of all sea states ω as

$$N = \omega \cdot T \quad (5)$$

Substitution of equation 4 and 5 into 3, results in a transient relationship for the change in crack depth as given by equation 6.

$$\int_{a_o}^{a_t} \frac{da}{CY^m[\pi a]^{0.5-m}} = T\omega \cdot A^m \cdot \Gamma\left[1 + \frac{m}{B}\right] \quad (6)$$

The fatigue limit state function is defined in equation 7, where equation 6 is used to represent the resistance to fatigue and load applied. Uncertainties due to stress concentration factor SCF , numerical calculation of stress S , geometric factor, and loading conditions F were also added to the fatigue limit state function.

$$g(X) = \int_{a_o}^{a_t} \frac{da}{C \cdot y \cdot Y^m[\pi \cdot a]^{0.5-m}} - T \cdot \omega \cdot A^m \cdot \Gamma\left[1 + \frac{m}{B}\right] \cdot SCF^m \cdot F^m \cdot S^m \quad (7)$$

2.1. Fatigue reliability index

Reliability index was calculated via AFOSM to determine the shortest distance between the origin and the failure surface of limit state function using Matlab. Random variables in the limit state function with non-Normal distribution were transformed to standard normal variables.

2.2. Update fatigue reliability index

The output from a SHM system mounted to an offshore fixed jacket infers the detection or non-detection of cracks in structure member(s). The detection or non-detection of crack was implemented to calculate an updated probability of failure P'_f using a conditional probability between the FLS and outcome from the SHM system O_i (Madsen et al. 1987, Moan & Song 2000, Shabakhty 2004)

$$P'_f = P[g(x) \leq 0 | O_i] \quad (8)$$

An outcome of crack detection by the SHM system is represented in the conditional probability based on time of occurrence T_o and crack size detected a_d as

$$O_{detection} = \int_{a_0}^{a_d} \frac{da}{C \cdot y \cdot Y^m [\pi \cdot a]^{0.5 \cdot m}} - T_o \cdot \omega \cdot A^m \cdot \Gamma \left[1 + \frac{m}{B} \right] \cdot SCF^m \cdot F^m \cdot S^m = 0 \quad (9)$$

For non-detection of crack, the SHM system detectable crack size a_{shm} , and time of occurrence are used to assess the SHM system outcome with

$$O_{non-detection} = \int_{a_0}^{a_{shm}} \frac{da}{C \cdot y \cdot Y^m [\pi \cdot a]^{0.5 \cdot m}} - T_o \cdot \omega \cdot A^m \cdot \Gamma \left[1 + \frac{m}{B} \right] \cdot SCF^m \cdot F^m \cdot S^m > 0 \quad (10)$$

Uncertainty in a_{shm} was characterized by its distribution function attained using model assisted probability of detection (MAPOD).

2.3. SHM damage detection and localization

Damage detection and localization was performed by analyzing modal frequencies generated from a fixed offshore platform using coordinate modal assurance criterion COMAC. COMAC was implemented to detect the variation in the mode shape for each component in the structure at multiple modes via equation 11.

$$COMAC_j = \frac{[\sum_{r=1}^L |\phi_{jr} \phi_{jr}^d|]^2}{\sum_{r=1}^L \phi_{jr} \sum_{r=1}^L \phi_{jr}^d} \quad (11)$$

Subscript r represents the number correlated modes attained from modal assurance criterion, j represents the mode shape location in the structure, while superscript d indicates mode shapes not attained from the structures reference state (Lieven & Ewins 1988).

COMAC values were normalized to attain a more sensitive damage index represented by equation 12, where $\mu COMAC_j$ and $\sigma COMAC_j$ are the mean and standard deviation of $COMAC_j$. The damage index signals the occurrence of damage by displaying a positive value that is relatively large, furthermore a value of damage index more than 2 indicates a potential location of damage (Cornwell et al. 1999).

$$Damage Index_j = \frac{\mu COMAC_j - COMAC_j}{\sigma COMAC_j} \quad (12)$$

In order to simulate crack size as damage, an equivalent stiffness k^e of a structure member with crack length a was modeled as

$$k^e = \frac{E \cdot I \cdot I_c}{0.125 \cdot \pi \cdot t \cdot D^3} \quad (13)$$

in which E represents Young's modulus, D is the average diameter between the pipe outside diameter, and inside diameter, t symbolizes the pipe wall thickness, I and I_c are the area moment of inertia of the undamaged and cracked structure member respectively (Lee et al. 2003). Crack depth was attained using crack aspect ratio of 0.2.

3. CONTRIBUTIONS OF SHM TO FATIGUE RELIABILITY INDEX

Numerical analysis was conducted via finite element model in SAPOS, to assess an example 3D offshore jack platform from a previous work by Karadeniz, using the same geometry, loading conditions, and material properties. The initial fatigue reliability was calculated using the random variables listed in table 1. The effect of SHM system on fatigue reliability was assessed by comparing the updated fatigue reliability index attained from two SHM systems with varying sensor density used to assess the integrity of the fixed offshore platform. Sensor densities for the two SHM systems are represented table 2 where the corresponding node locations on the structure are represented in figure 1. Three stages of damage were simulated in structure member 105 using three different crack lengths, where damage with a crack length of 2.5 mm occurred at year 5. The crack is then set to a length of 5 mm at year 8, and 10 mm at year 15. The stiffness of structural member 105 was reduced to an equivalent stiffness corresponding to the given crack length for each damage scenario.

Table 1: SHM system information

| SHM systems | # of sensor | Sensor Location | Element 105 crack length (mm) |
|-------------|-------------|--------------------------------|-------------------------------|
| SHM 1 | 8 | 24, 25, 26, 28, 34, 35, 37, 38 | 2.5, 5, 10 |
| SHM 2 | 8 | 6, 8, 9, 11, 30, 31, 32, 33 | 2.5, 5, 10 |

Table 2: Distribution of random variables

| Variable | Distribution | Mean | Coefficient of Variation |
|----------|--------------|---------------------|--------------------------|
| a_t | Normal | 8.0 | 0.04 |
| a_0 | Exponential | 0.11 ^a | |
| y | Normal | 1 | 0.1 ^a |
| $\ln C$ | Normal | -29.84 | 0.55* ^a |
| Y | Fixed | 1.0 | |
| m | Fixed | 3.1 ^a | |
| F | Lognormal | 1.0 | 0.1 ^a |
| S | Lognormal | 1.0 | 0.1 ^a |
| SCF | Lognormal | 2.50 | 0.15 ^a |
| ω | Fixed | 6.27e6 ^a | |
| A | Fixed | 10.9 | |
| B | Fixed | 1.22 ^a | |

* This value is a standard deviation. Superscript ^a indicates reference from Shabakhty 2004. Units are in N, and mm.

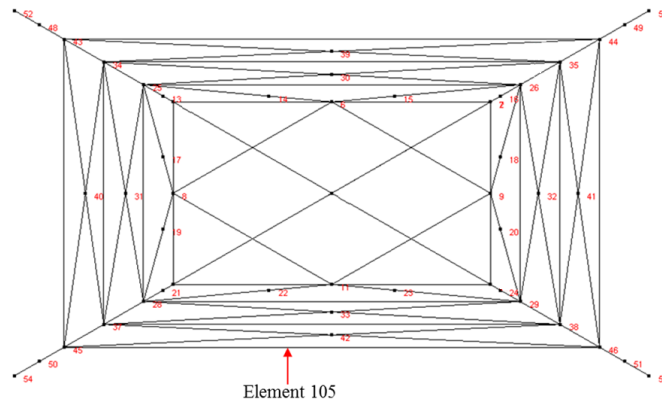


Figure 1. Nodal coordinates of the fixed offshore platform.

4. RESULTS AND DISCUSSIONS

The initial reliability of element 105 is displayed in figure 2 for a service life of 25 years, where β decreases with increasing service life. At a crack length of 2.5 mm, the structure modal frequencies were observed to display a faint reduction, and mode shapes from the first three modal frequencies were implemented to calculate COMAC, after which damage index was attained. Output for damage detection for SHM systems 1 indicates the detection of damage at node 38, where a relatively large positive value is indicated as seen in figure 3. For SHM system 2, there was no detection of damage as the positive values of damage index were below 2.

When a crack length of 5.0 mm was assessed, the maximum positive values for damage index increased in comparison to values attained from a 2.5 mm crack length, however a similar event observation was made by the SHM systems where SHM systems 1 and 2 indicated, damage and no damage respectively (See figure 3). A 10.0 mm crack length generated the detection of damage from SHM systems 1 and 2 where node 32 and 38 displays a value of damage index greater than 2 (See figure 4).

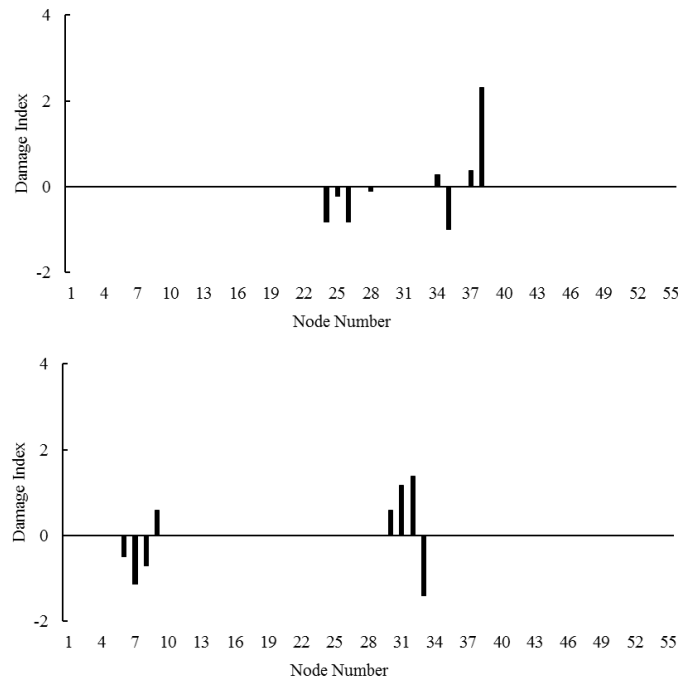


Figure 2. Damage index for the occurrence of a 2.5 mm crack length from (Top) SHM system 1 and (Bottom) SHM system 2.

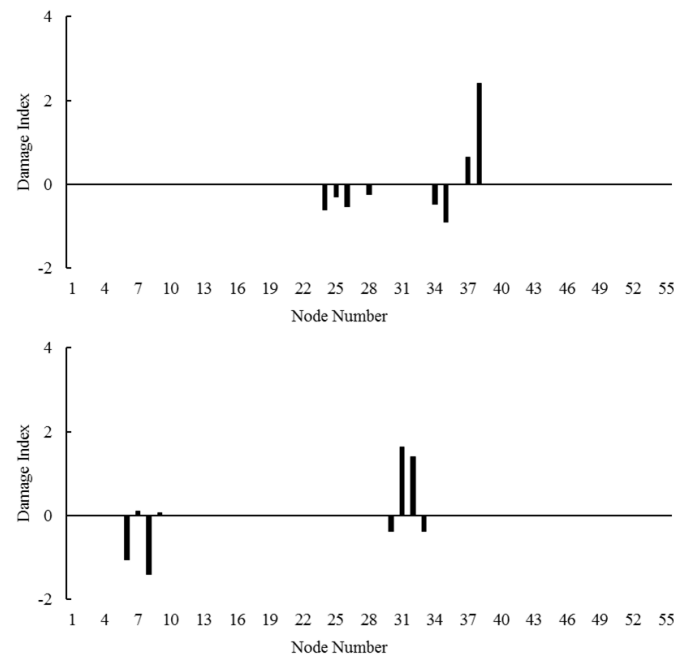


Figure 3. Damage index for the occurrence of a 5 mm crack length from (Top) SHM system 1 and (Bottom) SHM system 2.

The updated reliability index for structure member 105 using SHM system 2 in the presence of a 2.5mm long crack that occurred in year 5, and a 5 mm crack length that occurred in year 8 displays a similar reliability index pattern, where the updated reliability index is greater than the initial reliability index at all times through its operational life of 25 years. At year 15 where a 10 mm crack occurs, the updated reliability index displays values larger than the initial calculated reliability index for service time close to

the detection time. This trend was observed to change as the service time increased, where the updated reliability index dropped below the initial reliability index calculated.

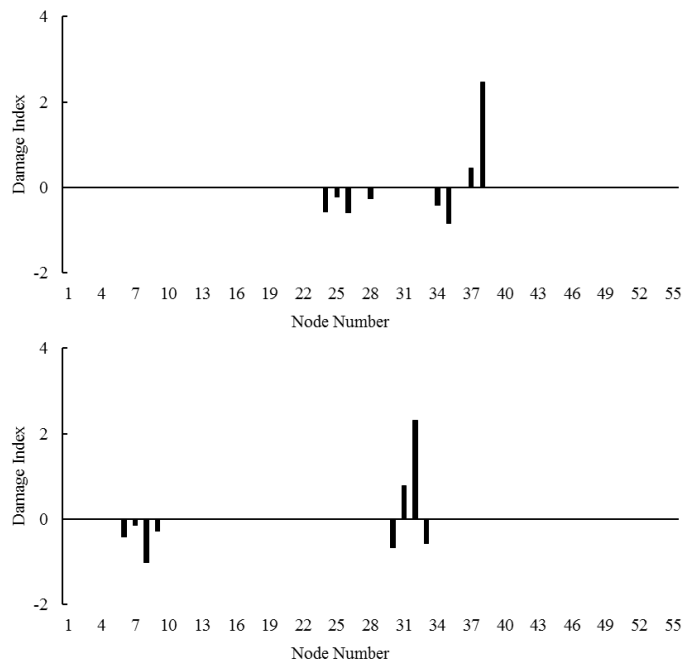


Figure 4. Damage index for the occurrence of a 10 mm crack length from (Top) SHM system 1 and (Bottom) SHM system 2.

Reliability index attained from SHM system 1 over the operational life structure member 105, in the presence of a 2.5 mm, 5mm, and 10mm crack length occurring at year 5, 8, and 15 respectively displayed similar trends. This trend was also identical to the reliability index attained from SHM system 2 in the presence of a 10 mm crack length.

The false negative outcomes from SHM system 2 for crack lengths of 2.5 - 5 mm, generates an overestimation of reliability index for member 105, when compared to SHM system 1. However at a crack length of 10 mm both SHM systems display similar reliability index.

5. CONCLUSION

In this study, the impact of SHM systems on the calculated reliability for an offshore platform structure member was assessed using fracture mechanics to model a fatigue limit state function. The initial structural member reliability index was calculated via first order reliability method, where uncertainties from initial crack length, final crack length, detectable crack length, geometric factor, fatigue material properties, force calculations, stress calculations, and stress concentration factors were characterized using their respective distributions. Two SHM systems with varying spatial locations were used to assess an example offshore platform system modeled using finite element analysis. Damage detection was conducted by means of coordinate modal assurance criterion, using 8 nodes to attain spatially incomplete modal data. Damage was simulated in the structure member using a partially connected beam with a spring, where the spring stiffness was modified to correlate with a given crack length. An outcome from the each SHM system was used to update the reliability index of the structure member for a given type of damage.

The results for updated fatigue reliability index showed that false negative outcomes from the SHM system generated updated reliability index that were above the initial reliability index as seen when SHM system 2 assessed the damaged member with crack length of 2.5 mm and 10 mm. Results from SHM system 1, indicated an accurate detection of 2.5mm and 10 mm crack length, the observed reliability index was observed to decline beneath the initial reliability index as the time in service approached the service life.

The contributions of SHM system to monitored structures reliability, in this work was based on the numerical detection of fatigue damage, however the additional forms of damage can be assessed by changing the limit state function. Furthermore, the damage detection technique used in this study does not specify the member where damage occurs, its only specifies the position of the spatially incomplete mode shape suspected of damage. Future work will assess the contribution of SHM systems to damage detection and localization in offshore platforms using numerical and experimental analysis.

REFERENCES

- [1] Cornwell, P., Doebling, S.W. and Farrar, C.R., 1999. Application of the strain energy damage detection method to plate-like structures. *Journal of Sound and Vibration*, 224(2), pp.359-374
- [2] Doebling, S.W., Farrar, C.R., Prime, M.B. and Shevitz, D.W., 1996. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review (No. LA--13070-MS). Los Alamos National Lab., NM (United States).
- [3] Karadeniz, H., 2001. Uncertainty modeling in the fatigue reliability calculation of offshore structures. *Reliability Engineering & System Safety*, 74(3), pp.323-335.
- [4] Lee, J.W., Kim, S.R. and Huh, Y.C., 2014. Pipe crack identification based on the energy method and committee of neural networks. *International Journal of Steel Structures*, 14(2), pp.345-354.
- [5] Lieven, N.A.J. and Ewins, D.J., 1988. Spatial correlation of mode shapes, the coordinate modal assurance criterion (COMAC). In *Proceedings of the sixth international modal analysis conference* (Vol. 1, pp. 690-695).
- [6] Madsen, H. O., Skjong, R., and Kirkemo, F, 1987. Probabilistic fatigue analysis of offshore structures--reliability updating through inspection results. In *International Symposium on Integrity of Offshore Structures (IOS'87)*, 3rd.
- [7] Moan T., and Song R, 1997. Implication of Inspection Updating on System Fatigue Reliability of Offshore Structures. *Journal of Offshore Mechanics and Arctic Engineering (OMAE)*, August, Vol.122, pp. 173-180
- [8] Pillai, T.M. and Prasad, A.M, 2000. Fatigue reliability analysis in time domain for inspection strategy of fixed offshore structures. *Ocean engineering*, 27(2), pp.167-186
- [9] Shabakhty, N., 2004. Durable reliability of jack-up platforms. *The Impact of Fatigue, Fracture and Effect of Extreme Environmental Loads on the Structural Reliability*. TU Delft, Delft University of Technology.
- [10] Vugths, J.H. and Kinra, R.K. 1976. Probabilistic fatigue analysis of fixed offshore structures. In *Offshore Technology Conference*. Offshore Technology Conference.

Experimental and Simulation Study on Impact Faults and Vibration Characteristics of Gear Box

Hui Zhipeng¹, Guan Tao¹, Wang Chen², Zhu Zhenqiao²

¹ China Marine Development and Research Center (CMDRC), Beijing, China

² Dianosis and Self-recovery Engineering Center, Beijing University of Chemical Technology, Beijing, China

ABSTRACT

The finite element analysis and experimental measurement are used to study the vibration characteristics of gearbox impact faults. Parametric models of gearbox gear train are built through three-dimensional modeling software, and the finite element software transient dynamics module is used to analyze the vibration acceleration response of the gearbox under different conditions of impact load. Compare the actual acceleration response which is collected from fault simulation experiment with the finite element simulation results to verify the accuracy of the parametric models. Meantime, the vibration response of the gear box impact failure is analyzed, and the sensor layout optimization scheme is proposed, which reduces the blindness of vibration measuring point arrangement and reduces the necessary channels for the monitoring and diagnosis. As a result, the cost of monitoring could be reduced.

Keywords: Gearbox, Modelling analysis and optimisation, Transient dynamics analysis, Monitoring and diagnosis

Corresponding author: ¹Hui Zhipeng (email: huizhipeng2017@163.com).

1. INTRODUCTION

Gearbox is an indispensable power transmission component for mechanical equipment. It is mainly used to change the speed and transmit power. It is widely used in wind power generation, aviation, ship, Metallurgical, petrochemical, mining, lifting and transportation industries such as mechanical transmission system, for it has the advantages of large transmission torque, high transmission precision and compact structure [1]. However, due to the gear box parts such as bearings, gears and shaft processing technology is more complex, high precision assembly, and often run at high speed continuously in harsh environment, parts failures appear as a result of wear failure, representing impact failure, in this condition machinery and equipment can't operate properly, may cause production disruption, economic losses even casualties[2].

In order to detect the failure of the gear box and reduce the occurrence of the accident, it is necessary to study the vibration characteristics of the gear box impact fault, analyze the vibration response [3], propose the sensor layout optimization scheme, reduce the blindness of the vibration measuring point arrangement, improve the accuracy of vibration detection, while reducing the monitoring and diagnosis of resource inputs.

Therefore, this paper adopts the method of finite element analysis and experimental measurement to study the vibration characteristics of gear box impact failure. The parametric model of wind turbine gearbox train is established by using 3D modeling software. The finite element software transient dynamics module is used to analyze the response of vibration compensation under different load conditions. The analysis theory uses the impact between the meshing surfaces as the basis for interpreting the dynamic excitation and response of the gear. The gear system is simplified to a simple single degree of freedom system and the impact load is applied. The dynamic response of the gear system is expressed by the dynamic response of

the gear system characteristic. Through the fault simulation experiment of the gear box, the actual acceleration response is measured and compared with the finite element simulation results to verify the accuracy of the parametric model. At the same time, the vibration response of gearbox impact failure is analyzed, and the optimization scheme of sensor layout is put forward, which provides a reference for the diagnosis of gear box impact fault.

2. PARAMETRIC MODELING OF GEARBOX

The structural parameters of the gearbox train system directly determine the transmission characteristics of the gearbox, so the accuracy of the established wheel train model has important influence on the later finite element simulation analysis. According to the structure of the wind turbine gearbox shown in Fig. 4.1, the parametric modeling of the gearbox is roughly as shown in figure 1

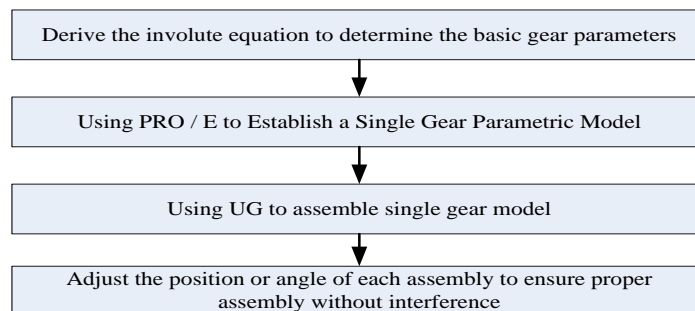


Figure 1. The main flow chart of modeling

In the assembly process, the most important thing is how to plan the wheel train, fixed shaft train to achieve the correct assembly, to avoid interference, the planetary gear assembly meets the corresponding assembly conditions. Finally, through the above-mentioned parametric modeling process to get the model shown in figure 2.

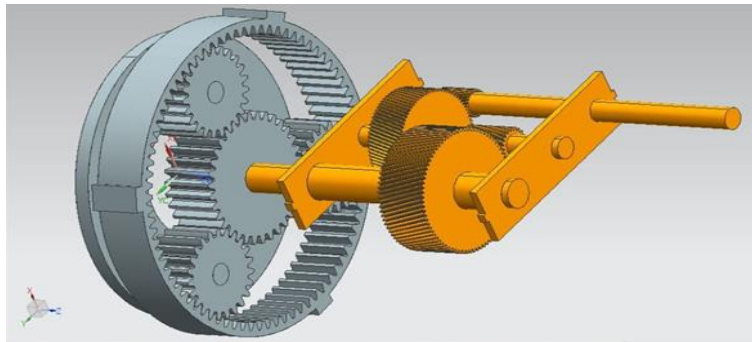


Figure 2. Gearbox train model

3. TRANSIENT DYNAMICS ANALYSIS OF GEAR BOX BASED ON ANSYS WORKBENCH

3.1. Transient Dynamics Theory

Transient dynamics analysis is a method used to determine the structural dynamic response of any load with time. The transient dynamical analysis gives the response of the structure to the time load. It is different from the rigid body dynamics analysis. The model of transient dynamics in ANSYS workbench can be rigid

body. It can be flexible body and can be contact, friction, collision and other dynamic analysis. Using the transient dynamics analysis can determine the displacement, stress, strain, velocity and acceleration of the structure with time under the static combination of static load, transient load and harmonic load.

3.2. Transient Dynamics Analysis of Gear Box

3.2.1. Gearbox model contact pair and motion pair setting

In the ANSYS workbench software the relationship between the two surfaces must be established to prevent the mutual penetration in the analysis, called the mandatory contact coordination. There are five different contact types, Bonded, No Separation, Frictionless, Rough, and Frictional. The contact involved in the study is a frictional contact and in view of the fact that the impact load may cause relatively large deformation of the contact unit. Thus, Enhanced Augmented Lagrange are used here, which adds additional control, can automatically reduce penetration.

The gearbox train model needs to define the contact pairs between the three planetary wheels and the meshing ring and the sun gear as well as the contact between the input shaft and the output shaft and the countershaft gear, as shown in figure 3 on the left.

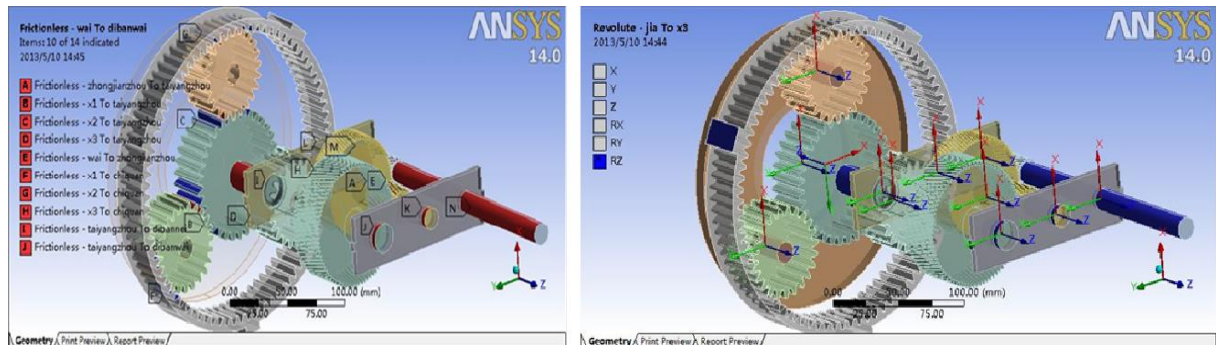


Figure 3. gearbox train contact and movement pair

Define the respective rotation pairs of each gear, including the rotation of the three planetary wheels and the pinion shaft on the planet carrier, the rotation between the sun wheel and the earth, the fixed pair between the ring gear and the earth, and the rotation between the three axes and the two supports fixed wheel train is shown in figure 3 on the right.

3.2.2. Gearbox model meshing

In the Method option, select the hexahedral dominant meshing method. This division method first generates a quadrilateral dominant surface mesh, and then obtains the hexahedron, and finally fills the pyramid and tetrahedral elements as needed. The hexahedral mesh can generate fewer elements than the tetrahedral lattice, accelerate the convergence of the solution, improve the accuracy of the analysis, and reduce the numerical error. Therefore, the hexahedral element is used as the dominant method.

Divide the grid as shown in figure 4.

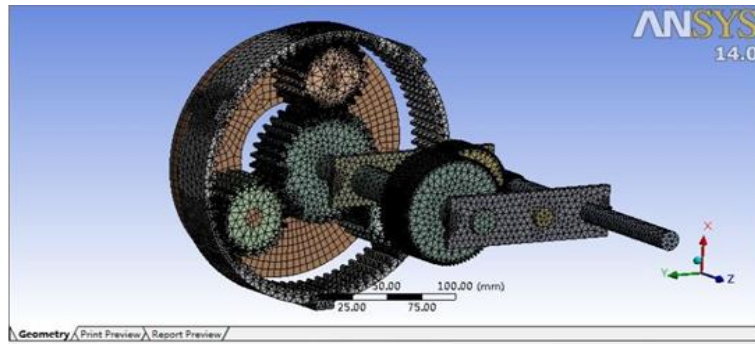


Figure 4. Gear mesh mode

3.2.3. Constraints and the application of loads

According to the most likely occurrence of the impact of the fault location, that is, the planetary wheel in the sun wheel and the planetary wheel meshing, the meshing between the large gear on the countershaft and the output shaft. Then the impact load is applied at these positions, and the outer surfaces of the outer teeth of the outer ring and the outer surfaces of the two outer surfaces of the fixed shaft train and the fixed constraints between the earth, to simulate the fault conditions. The vibration acceleration response is extracted from the calculated results and the position of the vibration acceleration response is determined. The acceleration sensor is arranged to optimize the sensor layout.

Therefore, the application of the finite element model load of the gearbox train is divided into the following two groups to simulate the impact vibration fault at the meshing tooth surface.

(1) Apply a pair of interaction forces at the meshing tooth surface of the sun gear with the planet wheel as shown in figure 5 on the left. An angular velocity load is applied to the center of the sun gear and counterclockwise, with a size of 3.0 rad/s ; a pair of intersecting forces perpendicular to the meshing surface are applied at one of the meshing teeth of the planetary gear and the ring gear in the opposite direction. The time of transient simulation is very short, about 0.01s , which are used to match the impact between the teeth of the tooth surface load.

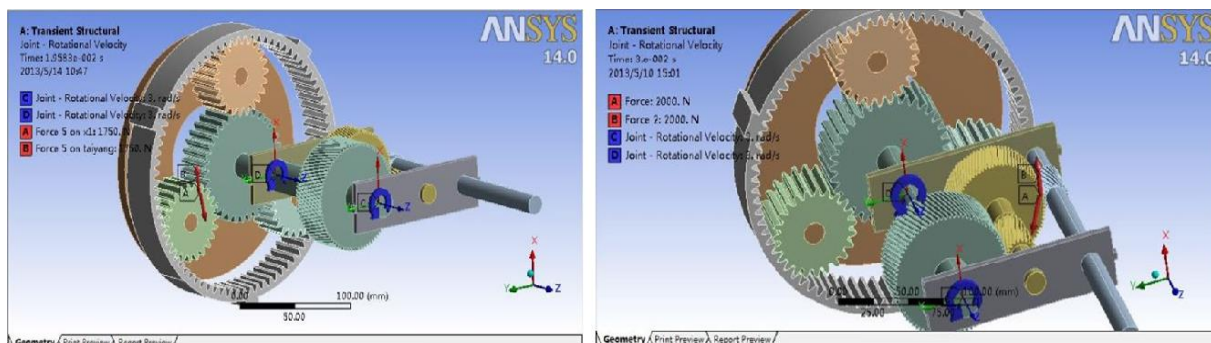


Figure 5. Loads are applied at the meshing surface

(2) Apply a pair of interaction forces at the meshing tooth surface of the helical gear of the fixed shaft output shaft helical gear and the intermediate shaft engaged therewith, as shown in figure 5 on the right.

The apply of the load is the same as in (1)

3.2.4. The results of transient dynamics analysis

Transient dynamics analysis requires the analysis and calculation of the finite element model of the gearbox train in the case of two different loads. The results are shown in figure 5

The instantaneous acceleration response of the sun wheel and the planetary wheel in the planetary gear train is as follows, and as shown in figure 5 on the left.

In the moment when the impact occurs, the vibration acceleration response of the area is mainly strong, and the vibration acceleration response is far from the central area, and the vibration acceleration response of the inner ring is more intense than that of the other region.

There is also a strong vibration acceleration response between the planetary gear and the inner ring gear. The more valuable result is that the acceleration axis of the output shaft of the fixed shaft train also has a vibration acceleration response, and the response to vibration acceleration is strong.

The acceleration response of the impingement of the meshing of the fixed shaft gear and the output shaft gear is as follows, and as shown in figure 6 on the right.

In the moment when the impact occurs, the two brackets of the fixed wheel train have vibration acceleration response.

In the same case, the vibration acceleration response is more obvious is the output shaft end position, which is consistent with the result of the analysis of the group (1), indicating that there is a certain regularity.

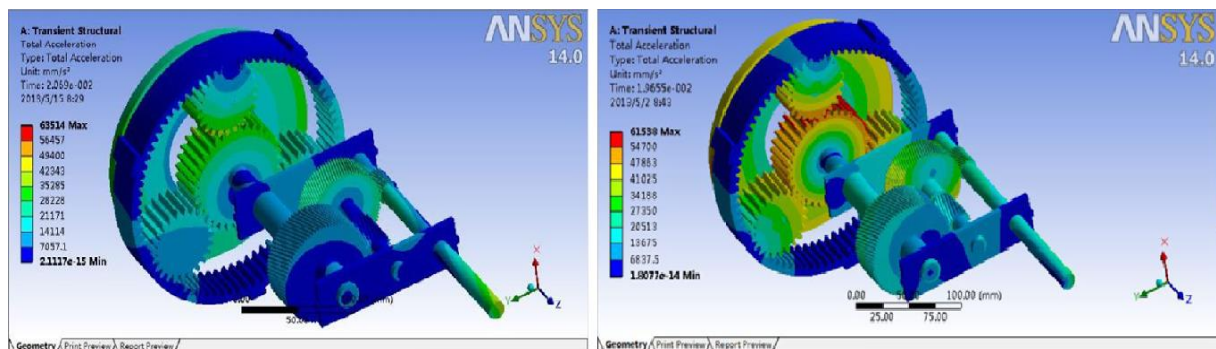


Figure 6. The results of transient dynamics analysis

In summary, the gearbox wheel train in the transmission process, when the impact of the planetary gear train vibration occurs, the outside of the ring gear observed in the more obvious vibration acceleration response; when the impact of the fixed gear train vibration occurs, a more obvious vibration acceleration response can be observed at the output shaft end of the fixed shaft train.

The more important conclusion is that the vibration acceleration response can be transmitted from the inside of the planetary gear train to the outside of the planetary gear train when the impact vibration occurs inside the planetary gear train, even the axis of the output shaft of the fixed shaft train, and the degree of vibration acceleration response can meet the acceleration range of the acceleration sensor.

4. EXPERIMENTAL MEASUREMENT OF VIBRATION ACCELERATION RESPONSE OF GEARBOX

4.1. Gearbox structure and basic parameters

The gearbox used in the experiment consisted of planetary gear train and fixed shaft train, with 8 teeth. The planetary gear train consists of three planetary wheels, the sun wheel and the fixed wheel train connected to the output of the sun wheel as a fixed wheel train input. The number of gears for each gear train is the same as that of the gearbox train model used in the analysis of the transient dynamics in Chapter 3, and the arrangement is exactly the same, as shown in figure 7.

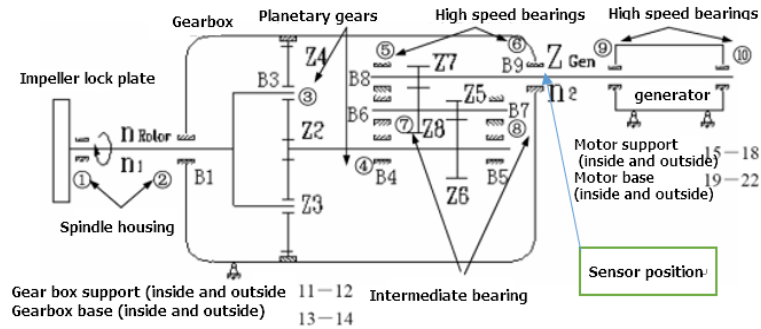


Figure 7. Gearbox structure and sensor position

Gearbox of the planetary gear train, $z_2=23$, $z_3=34$, $z_4=80$, when the motor input speed of 22 r/min, the gearbox axis frequency and the gear on which the meshing frequency see table 1.

Table 1: The frequency of each axis and the meshing frequency of the gears on it

| | Frequency (Hz) | Number of teeth | Engagement frequency (Hz) |
|--------------------|----------------|-----------------|---------------------------|
| Planetary frame | 0.36 | — | — |
| Planetary wheel | 2.36 | 23 | 54.28 |
| Sun axle | 1.65 | 34, 80 | 54.71, 128.80 |
| Intermediate level | 6.25 | 21, 81 | 129.25, 506.25 |
| High speed | 25.22 | 20 | 504.40 |

4.2. Gear fault presetting and sensor arrangement

The impact failure of the sun gear (the input shaft of the fixed wheel train) and the intermediate stage gear is simulated by machining the defect on the intermediate gear. The fault is shown in figure 8.

The acceleration sensor arrangement is shown in figure 7. It is verified by the results of the transient dynamics analysis, taking into account the rationality of the sensor mounting position in the gearbox. Therefore, the fault simulation experiment, in the gearbox output shaft axis horizontal and vertical direction of the layout of the acceleration sensor to measure the different parts of the gear box within the impact of vibration at the output shaft axis of the acceleration response to verify the transient dynamics Analyze the results.



Figure 8. Gear fault preset

5. DATA ANALYSIS AND RESULTS

Measure the vibration acceleration signal, use the online monitoring system software acquisition and analysis, the time domain processing to obtain vibration acceleration waveform, Fourier transform acceleration amplitude spectrum, as shown in figure 9.

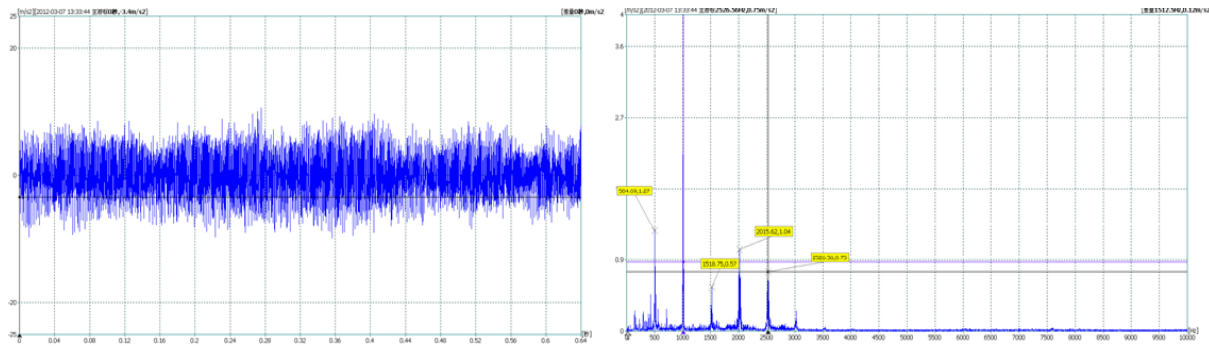


Figure 9. Vibration acceleration waveform and Spectrum

Waveforms are more obvious modulation phenomenon, the spectrum of high-frequency dominant, vibration. The spectrum is refined, as shown in figure 10.

In the refinement spectrum, the 129Hz frequency and its frequency doubling frequency 260Hz is dominated by the frequency component, and has a sideband, which is consistent with the gear engagement frequency (see table 1).

When the meshing point enters the defect of the gear, the gear will produce an impulse pulse, and the pulse signal can be decomposed into many sine quantities, so a series of side frequencies centered on the meshing frequency are formed on the spectrum, and the number is more, the amplitude is lower, the distribution is more uniform flat, forming the sideband [4]. And the width of the sideband directly reflects the frequency of the axis of the other gear engaged with the gear determined at the engagement frequency. In this way, the impact of the failure of the gear pair can be accurately positioned [5].

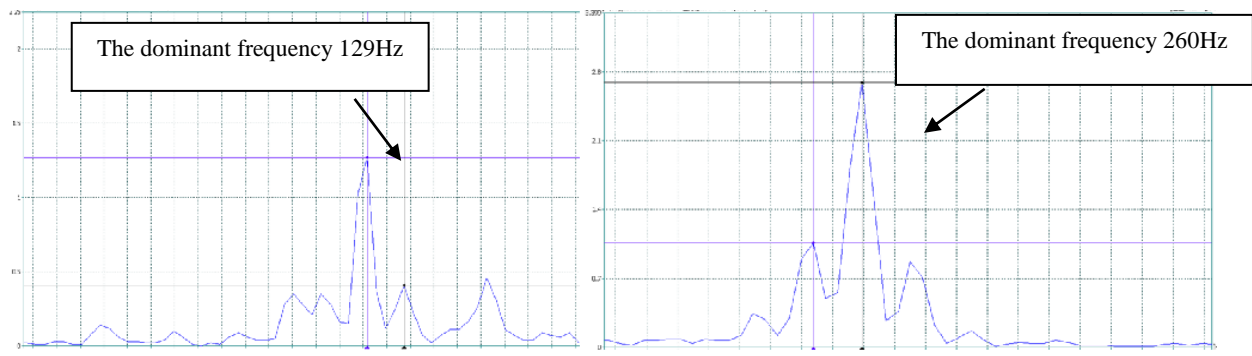


Figure 10. Refinement spectrum

From the experimental results, it is possible to effectively measure the vibration acceleration response by arranging the acceleration sensor on the output shaft end. This is consistent with the results of the transient dynamical analysis of ANSYS workbench in Chapter 3, and by analyzing the vibration acceleration signal, Accurate positioning of the location and type of shock vibration occurred, indicating that the establishment of gear box finite element model to analyze the impact of vibration response characteristics of the method is feasible, can follow the gear box measuring point layout optimization to provide guidance.

6. CONCLUSION

The transient dynamic analysis of the model was carried out by ANSYS workbench software and the vibration characteristics of the impact fault were studied. The response of vibration acceleration for finite element model under fault condition were obtained.

From the results of multiple sets of analysis and calculation we can see loading impact at different position, the output end of the fixed shaft train had obvious vibration acceleration response. This showed the vibration acceleration response can be transmitted from the inside of the gearbox train to the outside of the fixed wheel train until it reaches the output shaft of the fixed wheel train Shaft end.

The vibration response experiment of gearbox tested the transient dynamic response analysis made by ANSYS workbench software. The results show that the simulation results are in accordance with the experimental results, indicating the acceleration sensor can be arranged at the shaft end of the output shaft of the gearbox to collect the vibration acceleration signal.

The acceleration signal is further analyzed according to the geometric parameters of the gearbox, the frequency of each axis and the frequency of the actual fault signal, combined with the corresponding fault diagnosis mechanism, to find out the exact location of the failure gears which cause the impact vibration, in this way an effective fault monitoring is completed.

7. CONCLUDING REMARKS

Finite element analysis, experimental measurement and the vibration characteristics of gearbox impact failure are studied. The results of transient dynamic analysis and experimental measurement are compared. Based on the vibration measurement of gear box impact fault, the sensor layout optimization scheme is proposed, avoiding the blindness of the traditional vibration measuring point layout, reducing the cost of monitoring and diagnosis, and improving the accuracy of monitoring and diagnosis.

REFERENCES

- [1] Gong Yanyi, Chen Cunchun, Wang Yongjie. Calculation and structural design of involute cylindrical gear strength [M]. Beijing: Mechanical Industry Press, 1986. (In Chinese)
- [2] HAMEED Z, HON Y S, CHO Y M, et al. Condition monitoring and fault detection of wind turbines and related algorithms: A review[J]. Renewable and Sustainable Energy Reviews, 2009, 13(1): 1-39.
- [3] INALPOLAT M, KAHRAMAN A. A theoretical and experimental investigation of modulation sidebands of planetary gear sets[J]. Journal of Sound and Vibration, 2009, 323(4) : 677-696.
- [4] DEMPSEY P J, LEWICKI D G, LE D D. Investigation of current methods to identify helicopter gear health[C]. // A aerospace Conference, March 3-10, 2007, Big Sky, MT, USA. 2007 : 1-13.
- [5] Shen Qinggen, Zheng Shuiying. Equipment fault diagnosis [M]. Beijing: Chemical Industry Press, 2009, 243-257. (In Chinese)

Improved Multiwavelet Denoising Using Local Sliding Window for Multi-Fault Detection of Rotating Machinery

Jing Yuan, Jianhua Zhao, Ying Wei, Yu Zhou

Shanghai Radio Equipment Institute, Shanghai 200090, PR China

ABSTRACT

The multiwavelet sliding window technique could overcome the overkill of universal thresholding and detect the weak symptoms. However, the traditional denoising method using a rotating speed is favorable to retain the dominating features in a revolution cycle and appropriate for single fault detection. Thus, the improved multiwavelet denoising using local sliding window is proposed for multi-fault detection in the paper. The local sliding window technique utilizes a fractional rotating speed to segment the multiwavelet decomposition signals and then the local vector thresholding is performed in each local sliding window so as to extract and identify the multiple fault features. The proposed method is applied to gearbox multi-fault detection of hot strip finishing mill to verify its effectiveness.

Keywords: Multiwavelet denoising, sliding window, multi-fault detection.

Corresponding author: Dr J. Yuan (yuanjing_802@163.com)

1. INTRODUCTION

The multi-faults of rotating machinery usually characterize as the superimposed or cross effect of fault features. Due to their complexity, it is significant and challenging to beforehand detect the multi-faults for operating security assurance. The traditional fault detection methods such as empirical mode decomposition and cyclostationary [1] are always sensitive and applicable for single fault, which may not extract the various and compound fault features at the same time resulting in the failure of multi-fault detection.

By inner product transform [2], multiwavelets which possess multiple scaling and wavelet functions could match and extract the various fault features. Meanwhile, they have the important signal processing properties which the scalar wavelets fail to [3]. Thus, multiwavelet methods are potentially suitable for multi-fault extraction and detection, in particular the multiwavelet denoising techniques by eliminating the noises and enhancing the symptoms. Vector thresholding, multiwavelet neighboring, sliding window denoising and so on have continued to be investigated and improved [4]. Thereinto, the multiwavelet sliding window could overcome the overkill of universal thresholding and have shown the outstanding performance in fault detection [5]. However, the traditional sliding window denoising using a rotating speed is favorable to retain the dominating features in a revolution cycle and appropriate for single fault detection. To overcome the drawback, the improved multiwavelet denoising using local sliding window is proposed for multi-fault detection. The local sliding window utilizes a fractional rotating speed to segment the multiwavelet decomposition signals and the local vector thresholding is performed in each local sliding window so as to extract and identify the multiple fault features. The proposed method is applied to gearbox multi-fault detection of hot strip finishing mill to verify its effectiveness.

2. IMPROVED MULTIWAVELET DENOISING USING LOCAL SLIDING WINDOW

2.1. Summary of Multiwavelets

Starting from a signal f , a vector-valued input $\{s_0\}$ can be achieved by oversampling representation [6] and then the decomposition and composition of multiwavelet transforms are

$$s_{j-1,n} = \sum_k \mathbf{H}_{k-2n} s_{j,k}, \quad d_{j-1,n} = \sum_k \mathbf{G}_{k-2n} s_{j,k} \quad (1)$$

$$s_{j,k} = \sum_n \mathbf{H}_{k-2n}^* s_{j-1,n} + \sum_n \mathbf{G}_{k-2n}^* d_{j-1,n} \quad (2)$$

where $\{s_j\}$ and $\{d_j\}$ are the approximation and detail coefficients at the j^{th} scale, $\{\mathbf{H}_k\}$ and $\{\mathbf{G}_k\}$ are the lowpass and highpass filters, and the symbol $*$ is the conjugate transpose.

2.2. Local Sliding Window

Due to the periodic operations, the faults of rotating machinery such as gearboxes are involved by the rotating speed. Especially, the multi-faults often regularly generate the corresponding multiple features in each revolution cycle. The dynamic information in every fractional rotating speed is potential to perfectly describe the operating status and focus on the respective fault feature. Thus, the local sliding window by the fractional rotating speed is proposed for multi-fault detection. Let r be the rotating speed and f_s be the sampling frequency, the local sliding window at the j^{th} scale for $\{d_{j,k}, k=1, \dots, n\}$ is defined by

$$w_{j,m} = \{d_{j,k}, k=1+m\omega, \dots, (m+1)\omega\}, m=0, \dots, n/\omega-1 \quad (3)$$

where $\omega = r_i \times \frac{f_s}{2^j}$ is the width of the local sliding window (the point of ω is integer for computing) and $r_i = \frac{r}{i}$ is the fractional rotating speed chosen according to the transcendental analysis and fault prejudgment.

2.3. Local Vector Thresholding

Vector thresholding denoising extended from scalar wavelets mainly contains hard thresholding and soft thresholding. Due to the shrinkage in the soft thresholding reducing to smooth the fault features, vector hard thresholding [7] is performed in the local sliding windows to improve the signal-to-noise ratio and enhance the multi-fault features. For $\{d_{j,k}\}$ from $w_{j,m}$, the new standard term is defined

$$x_{j,m}(k) = \mathbf{d}_{j,k}^T \mathbf{V}_{j,m}^{-1} \mathbf{d}_{j,k} \quad (4)$$

where $\mathbf{V}_{j,m}$ is the covariance matrix of $w_{j,m}$, obtained by a robust covariance estimation. Then the local vector hard thresholding is designed by

$$\tilde{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{d}_{j,k} & \text{if } x_{j,m}(k) \geq \lambda \\ \mathbf{0} & \text{otherwise} \end{cases} \quad (5)$$

where $\lambda = 2 \log \omega$ is the threshold of $w_{j,m}$. It should be pointed out that although λ is equal for the local sliding windows at the same scale (that is the same threshold), $x_{j,m}$ is different for each $w_{j,m}$. Thus, the local thresholding is actually acted in $x_{j,m}$, rather than λ . To sum up, the improved multiwavelet denoising using local sliding window for multi-fault detection is performed as follows.

- (1) Starting from f , $\{s_0\}$ is achieved by oversampling representation;
- (2) Perform multiwavelet decompositions by Eq.(1) with the j^{th} scale;
- (3) Determine r_i according to the transcendental analysis and fault prejudgment;
- (4) Segment $\{d_{j,k}\}$ using $w_{j,m}$ by Eq.(3);
- (5) Calculate $x_{j,m}$ by Eq.(4) and threshold by Eq.(5) in $w_{j,m}$;
- (6) Regroup $\{\tilde{\mathbf{d}}_{j,k}\}$ by the inverse of Eq.(3);
- (7) Perform multiwavelet reconstructions by Eq.(2) using $\{s_{j,k}\}$ and $\{\tilde{\mathbf{d}}_{j,k}\}$;

- (8) Output the denoised \tilde{f} by the postprocessing. If \tilde{f} is not ideal, return to step 3 and change r_i ;
- (9) Identify the multi-fault features from \tilde{f} according to mechanical fault mechanism.

3. EXPERIMENTAL VALIDATION

We apply the proposed method to gearbox multi-fault detection of hot strip finishing mill to verify its effectiveness. The gearbox of the tested hot strip finishing mill has a single stage with Z22/Z65. Velocity sensors were amounted on the outer case of the gearbox and closed to the input shaft. One vibration signal measured at a sampling frequency 2560 Hz is shown in figure 1(a). It can be seen that the periodic impulses denoted as the red triangles occur in the original signal. Moreover, a series of faint impulses denoted as the green triangles also could be found from the data, however disturbed by the heavy noise. The rotating frequencies of the pinion and large gear were calculated to be $f_1=4.5727$ Hz and $f_2=1.5477$ Hz according to the meshing frequency 100.6 Hz obtain by the spectrum of figure 1(b).

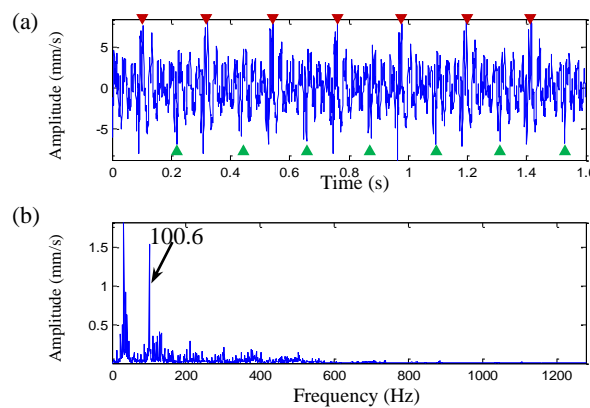


Figure 1. (a) Vibration signal; (b) Fourier spectrum.

The proposed method is introduced to analysis the measured signal. According to the interval time between the impulses denoted as the red and green triangles, one third of the rotating speed of the pinion is selected as the local sliding window, that is $r_i = 1/3 f_1$. Six-level multiwavelet decompositions using SA4 multiwavelets [8] are performed and denoised following the procedure in section 2.3. The analysis result by the proposed method is shown in figure 2. Biperiodic impulses of I_1 and I_2 are clearly seen from the result. The average period of I_1 and I_2 is both approximately 0.218 s, that is 4.5872 Hz corresponding to the rotating frequency of the pinion. Furthermore, the interval between I_1 and I_2 is approximately 0.085 s, close to 0.4 times of the rotating speed of the pinion. It could be indicted that two local faults appeared on the pinion, which had a distance close to 0.4 times of the circle. The contrastive results by the traditional multiwavelet sliding window denoising calculated by one rotating speed, multiwavelet vector thresholding and scalar wavelet (Db6) denoising using local sliding window depended on one third of the rotating speed are plotted from figures 3 to 5. Unfortunately, none of the contrastive methods could extract the intact impulses.

After the hot strip finishing mill was shut down and overhauled, two-site scuffing and wear faults were found on the pinion and the distance between them were 8 teeth, which were quite consistent with the conclusions.

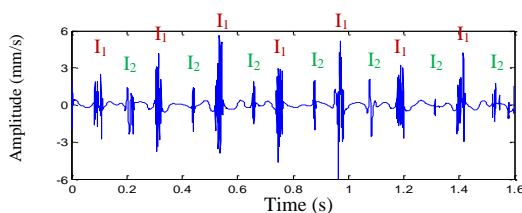


Figure 1. The analysis result by the proposed method.

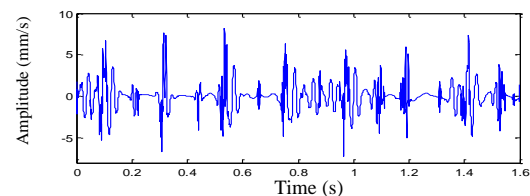


Figure 3. The analysis result by traditional multiwavelet sliding window denoising.

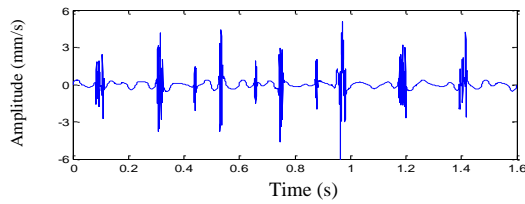


Figure 4. The analysis result by multiwavelet vector thresholding.

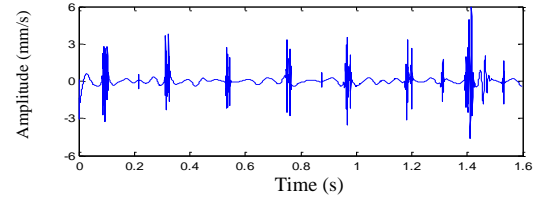


Figure 5. The analysis result by scalar wavelet denoising using local sliding window.

4. CONCLUSIONS

Due to the excellent properties of multiple basis functions, multiwavelet denoising methods are potential for multi-fault feature extraction. Especially, the multiwavelet sliding window technique could overcome the overkill of universal thresholding and detect the weak fault symptoms. However, the traditional sliding window denoising using a rotating speed is quite appropriate for single fault detection. Thus, the improved multiwavelet denoising using local sliding window is proposed for multi-fault detection. The local sliding window determined by the fractional rotating speed is proposed and vector hard thresholding is performed to extract the multi-fault features. The proposed method is applied to gearbox multi-fault detection of hot strip finishing mills, in comparison with the traditional multiwavelet and scalar wavelet denoising methods to verify its effectiveness. The results show that the method could provide a promising approach for multi-fault detection of rotating machinery.

ACKNOWLEDGMENTS

This work is sponsored by National Nature Science Foundation of China (No. 51405301), Shanghai Rising-Star Program (No. 16QB1403700), Shanghai Science and Technology Commission (No. 16111105900) and Shanghai Economy and Information Technology Commission (No. 160646).

REFERENCES

- [1] J. Antoni (2009) Cyclostationarity by examples. *Mechanical Systems and Signal Processing*. 23 (4): 987–1036.
- [2] Z.J. He, J. Yuan and Y.Y. Zi (2011) Inner product principle of mechanical fault diagnosis and applications. Science Publishing House, PR China. ISBN: 978-7-03-033119-9.
- [3] I. Daubechies (1992) Ten Lectures on Wavelets. CBMS-Conference Lecture Notes, SIMA Philadelphia.
- [4] H.L. Sun, Z.J. He, Y.Y. Zi, et al. (2014) Multiwavelet transform and its applications in mechanical fault diagnosis – A review. *Mech. Syst. Signal Process.* 70: 1-24.
- [5] J.Yuan, Z.J. He, Y.Y. Zi (2010) Gear fault detection using customized multiwavelet lifting schemes. *Mech. Syst. Signal Process.* 24: 1509-1528.
- [6] V. Strela. (1996) Multiwavelets: theory and application [D]. Cambridge: Massachusetts Institute of Technology.
- [7] V. Strela, N.Heller, G.Strang, et al. (1999) The application of multiwavelet filter banks to image processing. *IEEE Trans. On Image Processing.* 8(4): 548–563.
- [8] L. Shen, H.H. Tan, J.Y. Tham (2000) Symmetric-Antisymmetric orthonormal multiwavelets and related scalar wavelets. *Computational Harmonic Analysis.* 8: 258-279.

Session 2

Sensor Technology and Damage Detection 1

Session Chair Prof. Jing Lin

Monitoring of Water Contamination in Gearbox Lubricant Based on Vibration Analysis

(F. Khaldoun, F. Gu, A. Ball)

Weak Fault Diagnosis of Bearings Using a Cyclic Deconvolution Method by Combining SMHD with MEDA *(Y. Miao, M. Zhao, J. Lin, J. Zhao, J. Wang)*

Quantitative Detection of Steel Surface Cracks Based on Eddy Current Pulsed Thermography

(J. Ma, X. Xu, D. Zhen, H. Zhang, Z. Shi)

Cluster-based Thresholding of Phased Array Ultrasound for Anomaly Detection in Weld Inspection *(B. Cassels, L. Shark, S. J. Mein, A. Nixon, R. Turner)*

A Novel Two-Stage Approach to On-Site Condition Monitoring *(V. Leavers)*

Monitoring of Water Contamination in Gearbox Lubricant Based on Vibration Analysis

Khaldoon F. Brethee^{1,2}, Fengshou Gu¹, Andrew D. Ball¹

¹Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH.

²Engineering College, University of Anbar, Ramadi, Iraq.

ABSTRACT

To develop online condition monitoring of lubrication characteristics, this research will investigate the effect of lubricant water contamination on the gearbox vibration signatures. A theoretical analysis is conducted to understand the changes of lubricant properties caused by added water content. Then it predicts how the change can alter the tribological behaviour of gear mesh process that can critically lead to changes in vibrations of gears. Moreover, experimental studies have obtained consistent results which allows for an indication of the lubrication performance, showing different degrees of water contamination. Gearbox vibration signal was analysed using modulation signal bispectrum (MSB) analysis, which has high performance in selecting modulation components and suppressing random noise. More accurate and reliable results were produced using this method for assisting and diagnosing lubricant condition.

Keywords: gearbox lubrication; condition monitoring; water contamination; viscosity; vibration signal.

Corresponding author: Khaldoon F. Brethee (Khaldoon.Brethee@hud.ac.uk; khaldon77m@hotmail.com)

1. INTRODUCTION

The reliability, life time and availability of the most rotating machines such as gearbox depend largely on the lubrication condition. In modern industries, lubrication oil plays a critical part in condition maintenance of rotating machineries. Recently, health condition monitoring and prognostics of lubrication oil have become a significant topic among academia and industry [1, 2]. The purpose of most research is to provide early warning of machine failure and also extend the operating duration of lubrication oil.

Lubricant failure causes equipment failure and vice-versa, hence in many cases, lubricant failure is due to oil degradation by external contamination with water, dust particles, etc. [3]. Water is one of the most significant destructive contaminations to lubricants, which causes a degradation on its characteristics and excessive water contaminant leads to insufficient lubrication and subsequently to abrasive wear and corrosion as reported by Kuntner et al. [4]. Therefore, monitoring external contamination of lubrication is an important issue in a proactive maintenance program.

Vibration analysis is an effective technique that widely used for monitoring and diagnosis of various industrial applications. Despite the fact that many problems are caused by lubrication failure, very few works examine the possibility of employing vibration analysis for detecting lubrication statues [5-8]. It can be effective for performing the instant diagnosis of water contamination so as to prevent further damages to gearboxes running with water contents, as stated by Abusaad et al. [9]. However, vibration signal needs an effective signal processing method to extract transient signals and suppress random noise. MSB was developed based on the demodulation characteristics, which has the capability to suppress the noise background and extract useful features from the mechanical vibration signals [10, 11].

This study develop online condition monitoring of gearbox lubrication based on vibration signal analysis using an effective denoising method, as an approach to define a relation between the vibration signature and the lubrication condition. An experimental investigation was carried out using sequences of water contents dropped to gearbox lubrication under different operating conditions. MSB was applied to the

gearbox vibration signal as an approach in extracting effective diagnostic information based on the utilization of the modulation characteristics.

2. GEARBOX CONDITION MONITORING PROCESS

The aim of Lubricant condition monitoring is to ensure that lubricant has retained its physical characteristics, and it has not degraded by external contaminants. Water is one of contaminant present in lubricant that can cause corrosion and oxidation behaviour of lubricants and damage failure to the operating surfaces [12]. Vibration analysis is one of the most effective techniques for monitoring the health operation of machinery to avoid catastrophic failures of the system [13]. It is relatively an easy task for monitoring gearbox operating conditions [14]. The main steps of gearbox condition monitoring based on vibration signature analysis are shown in figure 1.

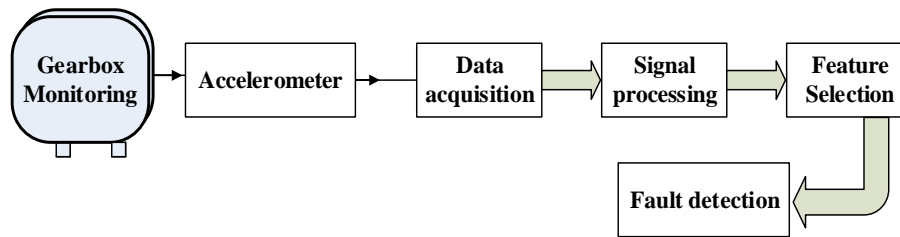


Figure 1. Step configuration of gearbox oil level condition monitoring

3. EXPERIMENTAL SETUP AND TEST PROCEDURE

To study the effect of water contamination on gearbox vibration signature, an experimental work was carried out by adding various amounts of water to the gearbox oil lube. The test rig comprised a back to back two-stage helical gearbox 13.1 kW, driven by a three-phase induction motor 15kW, and connected to a DC generator, in which different operating conditions can be applied to the system. An electromechanical accelerometer mounted to the gearbox housing was used to measure the gearbox vibration signals, as shown in figure 2.

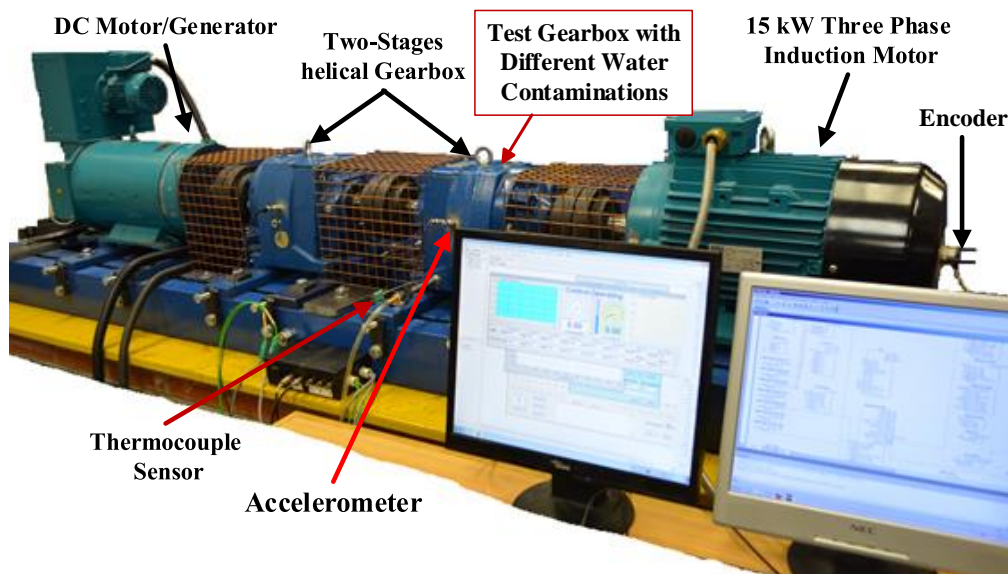


Figure 2. Test rig construction

A schematic structure of the tested gearbox (two stage helical gear box manufactured by David Brown Radicon Limited) is shown in figure 3. The oil lubrication of the gearbox is mineral oil of 320 EP with

recommended volume capacity of 2.6 Litres. Different quantities of water contaminations were simulated up to 120 kppm. The tested gearbox is a speed reducer with an overall reduction ratio of 3.678. Some gears are fully submerged in the oil and some of them are splashed. The rotational and meshing frequencies of the gearbox components can be verified by:

$$f_{r1} = \frac{\omega_1}{60}, \quad f_{r2} = f_{r1} \frac{Z_1}{Z_2}, \quad f_{r3} = f_{r1} \frac{Z_1}{Z_2} \frac{Z_3}{Z_4} \quad (1)$$

$$f_{m1} = f_{r1} Z_1, \quad f_{m2} = f_{r2} Z_3 \quad (2)$$

where, ω_1 is the input shaft speed in (rpm), Z_i refers to the gear teeth number, f_{ri} refers to the shaft frequencies and the mesh frequency of the first and second stage of the gearbox are denoted by f_{m1} and f_{m2} , respectively.

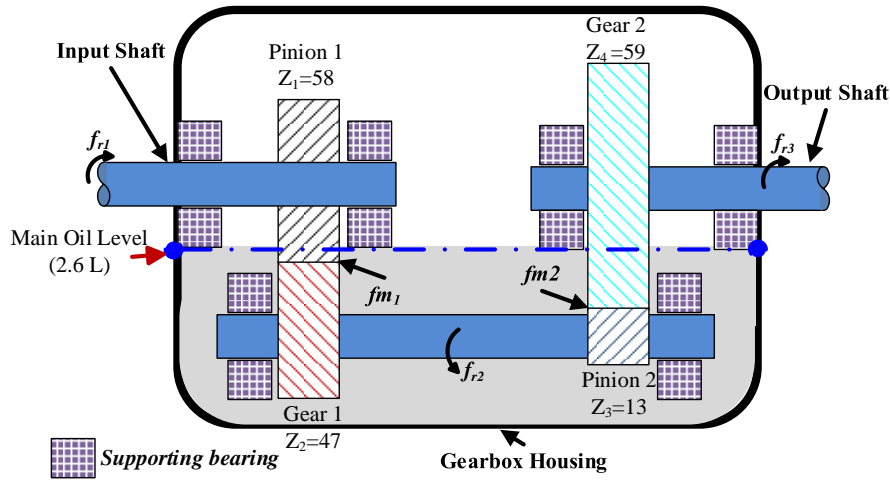


Figure 3. A schematic description of two-stage helical gearbox

4. EFFECT OF WATER CONTAMINATION

Water in oil mixture is the most destructive contaminant, which undermines the lubricant performance from forming an effective lubrication film between the contact components. It causes significant chemical and physical changes [2]. Hence, detection of water contamination has received significant attention to avoid catastrophic failures and increase the life cycle of equipment [9]. The typical acceptable water content for transmission oils are in the range of 1 to 2% (10 to 20 kppm) [15, 16]. Therefore, the test was simulated different values below and above this range, which allows a variety of different underlying measurements to be examined in a wide range for defining their corresponding detection performances.

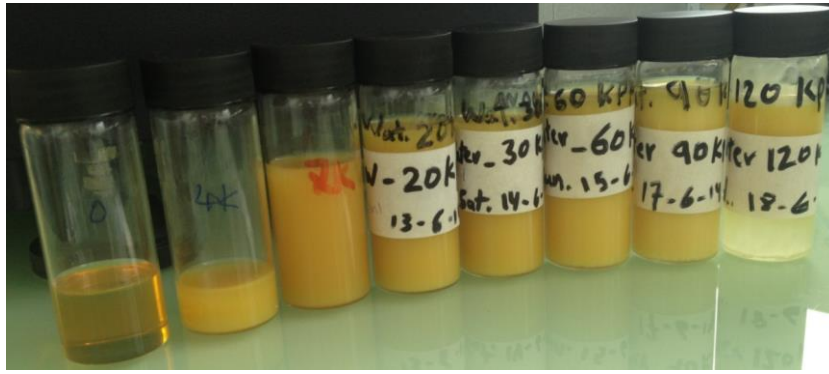


Figure 4. Gearbox oil samples with different water contaminations

As oil viscosity is the most important parameter of lubrication; a Kinexus pro+ rheometer was used to measure the oil viscosity and taken as references to represent the deterioration progress due to water contamination. Different MILLGEAR 320 EP (320 cSt at 40°C) samples were taken from gearbox housing just when the test was stopped, as shown in figure 4. It can be seen that the oil colour was changed due to water content and there is free water in the sample content of 120 kppm. This confirms that the water content was added effectively according to the test design.

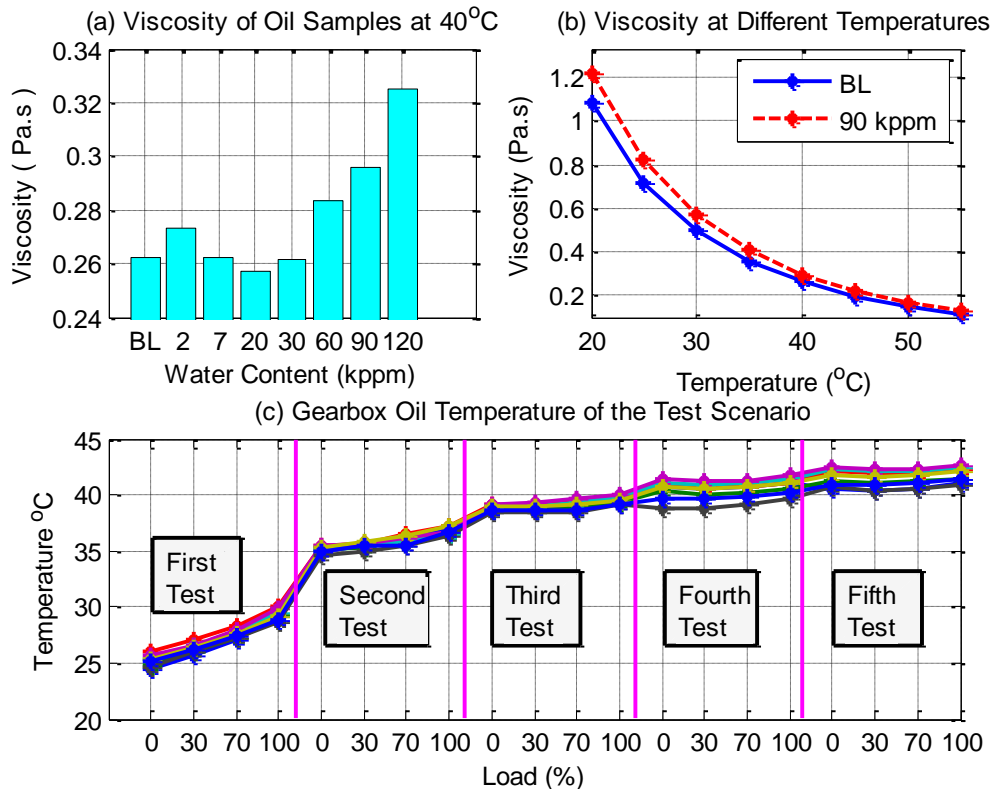


Figure 5. Viscosity of lube oil with different water contents at selected temperatures

Figure 5 (a) shows the representative viscosity values with different water contamination at the stable operating temperature of the gearbox. It can be seen that the viscosity is generally increased with the interaction extend of the water droplets. Moreover, temperature has a significant influence on the viscosity, i.e. the higher temperature corresponding to lower viscosity, see figure 5 (b). The oil temperature of the test scenario is shown in figure 5 (c), whereas the oil temperatures with different water contents are in the same range for all the repeated runs.

5. FRICTIONAL EXCITATIONS OF HELICAL GEARS

Friction force variation has large fluctuation along gear meshing line, which is the main cause of excitation in gears. The effects of friction on meshing gears include nonlinear dependence of friction on the sliding velocity, and energy dissipation [17, 18]. In general, frictional effects can be derived from elasto-hydrodynamic lubrication (EHL) and tribology theory, which is considered the dominant mode of lubrication accumulated with the gears meshing surfaces [19]. For successful operation mechanism of EHL, a numerical formula of friction was proposed by Xu et al. [20] to study the effect of system parameters on friction coefficient:

$$\mu = e^{f(SR, P_h, v_o, S)} P_h^{b_2} |SR|^{b_3} V_e^{b_6} v_o^{b_7} R^{b_8} \quad (3)$$

$$f(SR, P_h, v_o, S) = b_1 + b_4 |SR| P_h \log_{10}(v_o) + b_5 e^{-|SR| P_h \log_{10}(v_o)} + b_9 e^S \quad (4)$$

where SR is slide-to-roll ratio, P_h is the contact pressure, v_o is the dynamic viscosity of lubricant, S is the tooth surface roughness, V_e is the entering gear velocity, R is the combined radius of curvature and b_i are constants explained in the same reference. Since, viscosity is an important parameter for oil quality. It can generally increase due to lubricant oxidation and degradation or contamination with water, dust etc. Figure 6 shows the effects of lubricant viscosity variations between meshing gears on the time varying of friction coefficient with slide-to-roll ratio (SR) based on the previous numerical formula. It can be seen that, the time-varying friction coefficient is increased with increasing oil viscosity.

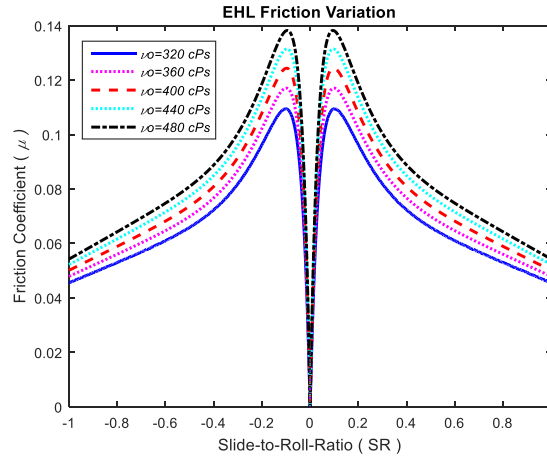


Figure 6. Time varying friction coefficient relationship with SR for EHL

6. RESULTS AND DISCUSSION

The main objective of this study is to investigate the effect of lubricant water contaminations on the vibration characteristics of the gearbox. The general behaviour of the gearbox vibration spectrum with different water contents under various operating conditions are shown in figure 7. Gear mesh frequency components (up to 3th harmonics) are also shown.

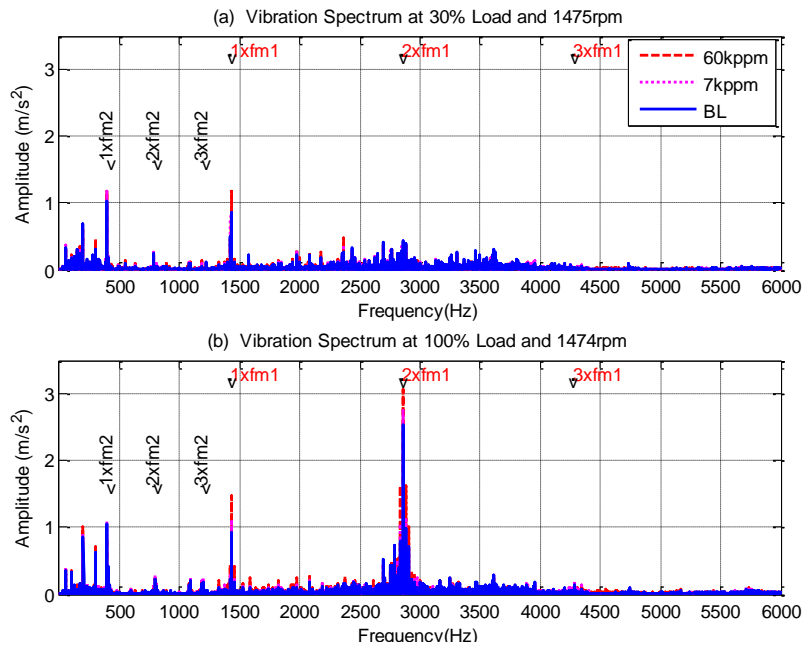


Figure 7. Vibration spectrum of the tested gearbox

6.1. Modulation signal bispectrum calculation (MSB)

For a reliable operating condition of machinery, various methods for fault detection and diagnosis have been developed. Vibration analysis and associated signal processing techniques are widely used as an approach to support condition monitoring and diagnostic procedures. To extract diagnostic information from modulation characteristics of fault frequency, MSB has the capability to decompose nonlinear modulation components and suppress random noise by detecting phase coupling in the modulation signal. The MSB is described by the correlation between the components of gear mesh frequency f_m and sidebands $f_m \pm f_r$ as [20]:

$$B_{MS}(f_m, f_r) = E \left\langle X(f_m - f_r) X(f_m + f_r) X^*(f_m) X^*(f_m) \right\rangle \quad (5)$$

where f_r is modulated frequency; f_m is the carrier frequency, $f_m + f_r$ and $f_m - f_r$ are modulation frequencies, $X^*(f)$ is the complex conjugate of the Fourier transform $X(f)$ of vibration signal $x(t)$; and $E[\]$ is the statistical expectation operator. The magnitude and phase of MSB can be expressed as:

$$A_{MS}(f_m, f_r) = E \left\langle \left| X(f_m - f_r) \right| \left| X(f_m + f_r) \right| \left| X^*(f_m) \right| \left| X^*(f_m) \right| \right\rangle \quad (6)$$

$$\phi_{MS}(f_m, f_r) = \phi(f_m - f_r) + \phi(f_m + f_r) - \phi(f_m) - \phi(f_m) \quad (7)$$

MSB take into account the lower sideband $f_m + f_r$ and the higher sideband $f_m - f_r$ simultaneously to measure the nonlinear effects of modulation signals. In the meantime, other components including random noise and interfering components that are not meet the phase relationship will be suppressed significantly. In addition, the coherence of MSB is used to estimate the influence of random components and to ensures the quality of MSB magnitude can be effective detector of modulation. The coherence can be defined by:

$$b_{MS}^2(f_m, f_r) = \frac{|B_{MS}(f_m, f_r)|^2}{E\langle X(f_m)X^*(f_m) \rangle E\langle |X(f_m + f_r)(f_m - f_r)|^2 \rangle} \quad (8)$$

The MSB coherence has boundary [0 1], whereas 1 means MSB magnitude is from true modulation effects and zero is mainly from random noise influences.

6.2. MSB characteristics at mesh frequency

Spectral amplitude variations are commonly used in the detection of abnormal change and diagnostic gearbox conditions. The vibration signal of the test gearbox was processed by MSB to suppress noise and irregular influences. As gear mesh frequency is more associated with gear dynamics, figure 8 shows the averaged amplitudes at the components of the first mesh frequency for the five repeated runs under different operating conditions. It can be seen that the vibration amplitudes at the f_{m1} harmonics are almost the same and there is not clear indication for the water contamination. This could be because this stage is operated at high speed with low load, and the mesh region is not fully submerged in the oil lube.

In the same way, the vibration amplitudes up to third harmonics of f_{m2} , which is fully submerged and influenced by low speed stage with high load transmission are shown in figure 9. It can be seen that the trends exhibit within more consistent behaviours for different loads and speeds, in which the third harmonic ($3*f_{m2}$) is significantly increased by the change of water contents and can be a good indicator to lubricant deterioration. However, there is a slight decrease in the amplitude of this harmonic under high speed and full load, due to the fact that under this condition frictional forces and oil squeezing are acting on each other depending on speed, load and temperature statuses.

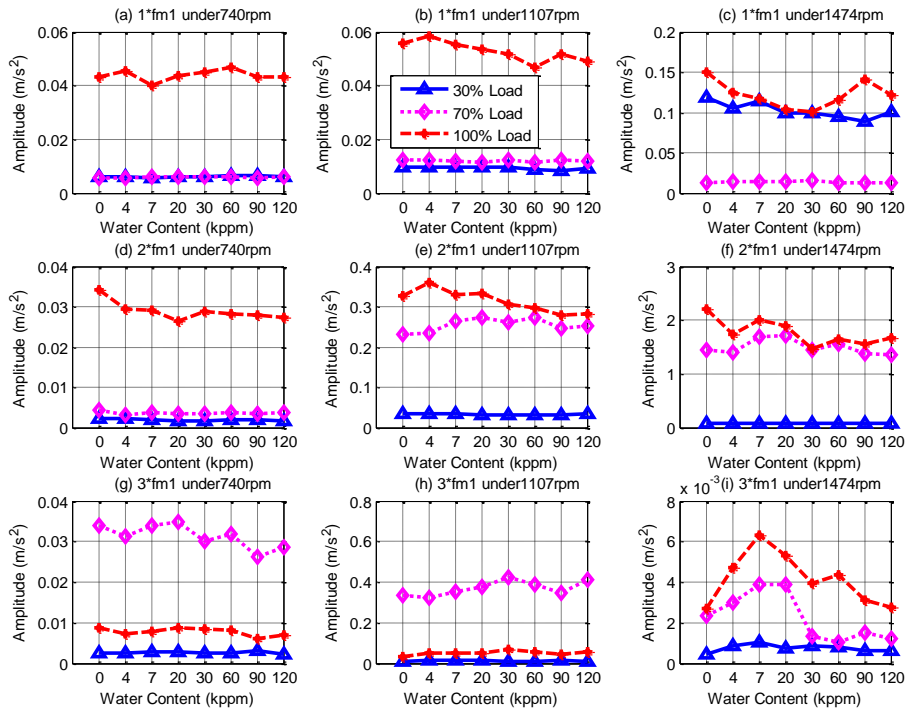


Figure 8. Average amplitudes of the first mesh frequency components under different operating conditions

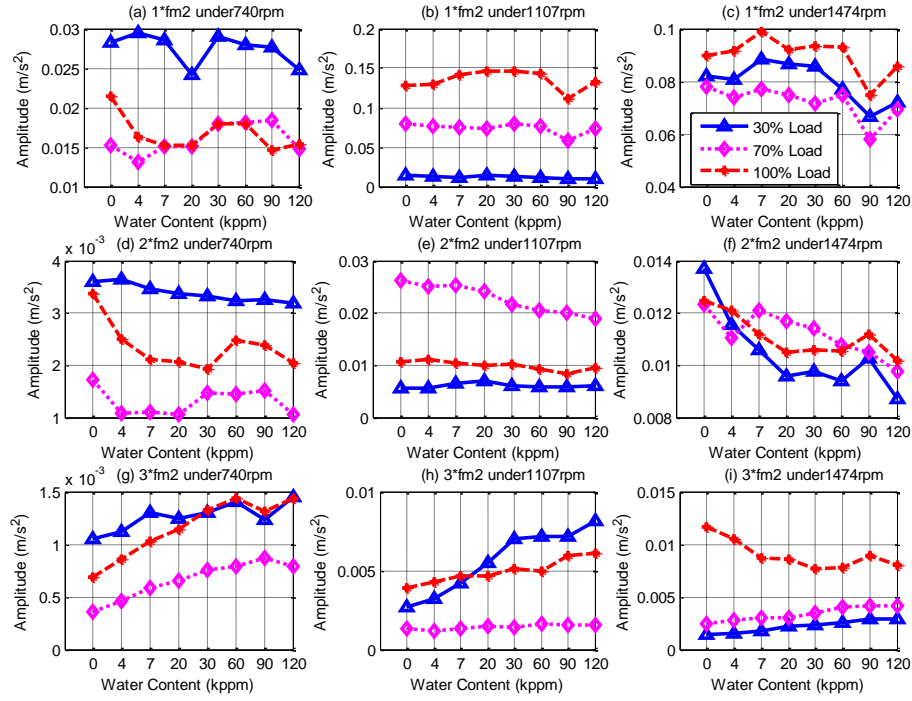


Figure 9. Average amplitudes of the second mesh frequency components under different operating conditions

6.3. MSB characteristics at sideband components

As the third harmonic ($3*f_{m2}$) exhibits more beneficial behaviour than the other harmonics, sideband frequencies around this component have been examined to extract the usefulness of these sidebands in diagnosing. Based on the gearbox geometry, the mesh frequency harmonics can be modulated with the components of shaft frequency at $n*f_{m2} \pm n*f_{ri}$, where $n=1, 2, 3, \dots$. Figure 10 shows the MSB result of these sidebands modulated with the $3*f_{m2}$ along with their coherence, in which the MSB magnitude can be considered as an effective detector of modulation at these sidebands.

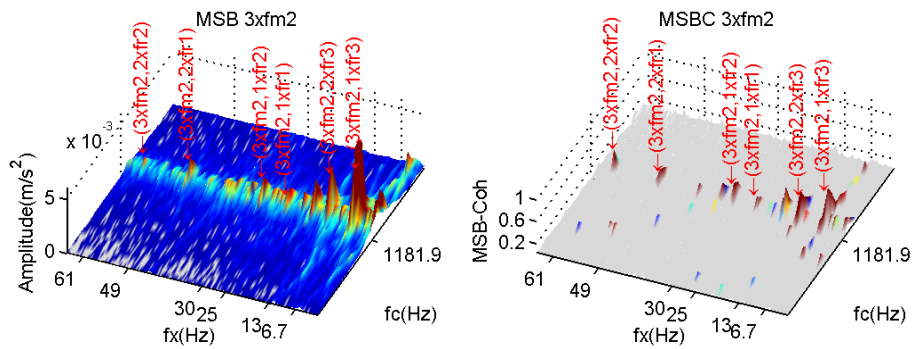


Figure 10. MSB results around $3xf_{m2}$ with sideband components at high speed

The average amplitudes for all the repeated runs at these sidebands are extracted from MSB results with different water contents. Figure 11 presents the average amplitudes at the two components of the input shaft ($1xf_{r1}$, $2xf_{r1}$) under different operating conditions. It can be seen that the amplitudes are generally increased with water content for the most working conditions.

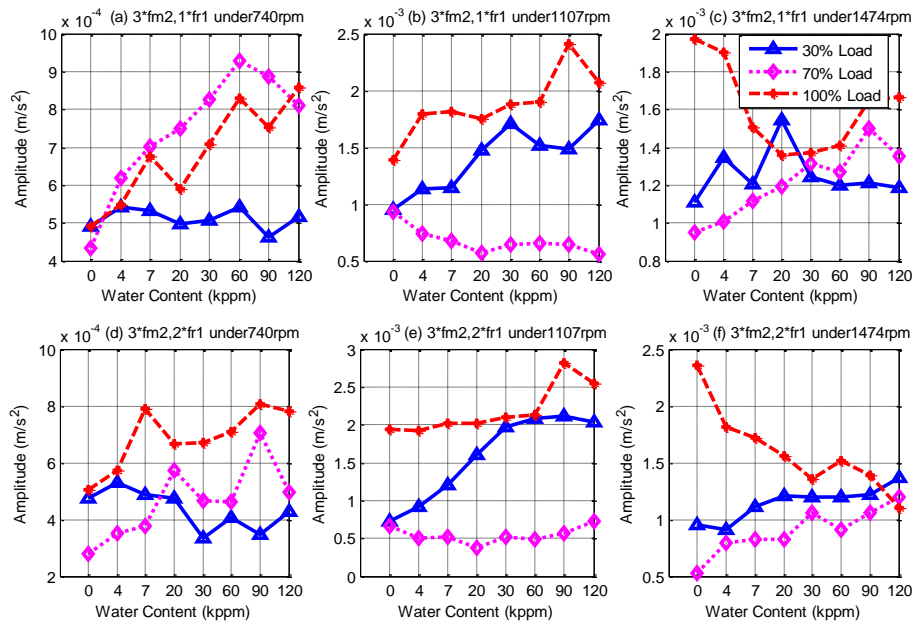


Figure 11. Average amplitudes of the input shaft components under different operating conditions

The average amplitudes of MSB results at the components of the intermediate shaft ($1xf_{r2}$, $2xf_{r2}$), which is the highest transmission shaft speed, are shown in figure 12. The amplitudes are also increased with water content, however for the high speed the trends are fluctuating. This may be because the effect of sliding friction between the engaged gears is decreased with increasing the running speed.

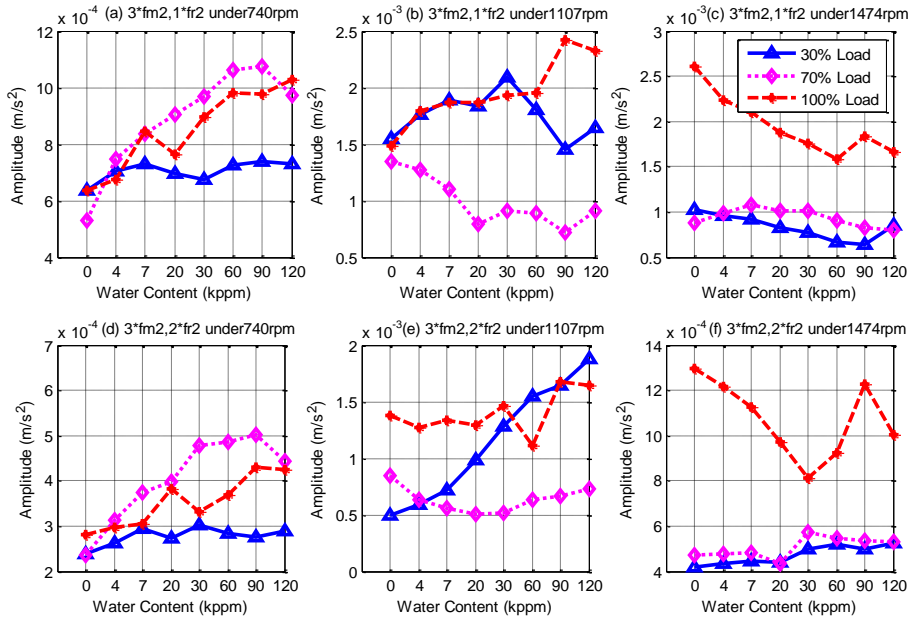


Figure 12. Average amplitudes of the second shaft components under different operating conditions

Similarly, the average amplitudes of MSB results at the components of the third shaft ($1xf_{r3}$, $2xf_{r3}$), which is the lowest transmission shaft speed, are shown in figure 13. It can be seen that the amplitudes behaviours are very close to the $3xf_{m2}$, which are generally increased with water content and high influence of modulation at these components under different operating conditions. In which this feature can be also considered as effective indicators to the lubrication oil degradation.

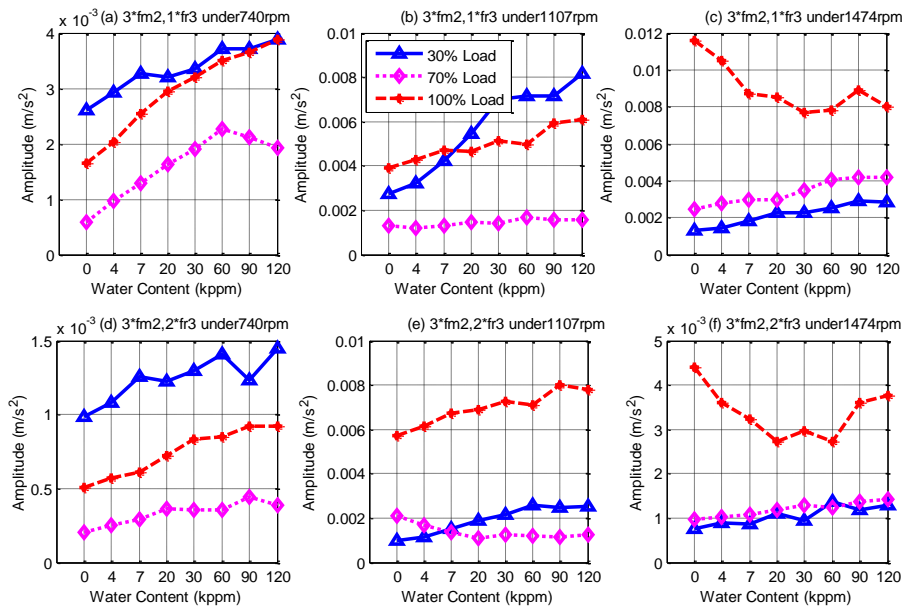


Figure 13. Average amplitudes of the third shaft components under different operating conditions

7. CONCLUSION

Vibration signature of gearbox shows complicated modulation of mesh frequency with different shaft frequencies, which cause inconsistent variation in spectral amplitudes and result in excessive difficulties to define a stable diagnostic feature. MSB analysis is an effective method to select effective modulation components and suppress random noise. This study examines vibration signature of two-stage helical gearbox with different water contaminations under different operating conditions, in order to investigate any measurable changes that can be correlated with the lubricant attenuation. The results show that the third harmonics of mesh frequency and its sidebands can produce a useful information for assisting and diagnosing lubricant condition, especially when the nonlinear modulation components decompose effectively. However, under high load and high speed it is difficult to use this features for a stable diagnostic.

ACKNOWLEDGMENT

The authors acknowledge the Higher Committee for Education Development in Iraq (HCED) and the University of Anbar for their supports.

REFERENCES

- [1] Zhu, J., *Online Industrial Lubrication Oil Health Condition Monitoring, Diagnosis and Prognostics*. 2013, PhD Thesis, University of Illinois at Chicago.
- [2] Zhu, J., D. He, and E. Bechhoefer, *Survey of lubrication oil condition monitoring, diagnostics, and prognostics techniques and systems*. Journal of Chemical Science and Technology, 2013. 2(3): p. 100-115.
- [3] Graça, B. and J. Seabra, *IMPROVING REALIBILITY OF WIND TURBINE GEARBOXES THROUGHT OIL ANALISIS*. 15th International Conference on Experimental Mechanics, ICEM15, 22-27 July 2012, Porto/Portugal, .
- [4] Kuntner, J., R. Chabicoovsky, and B. Jakobý. *Oil condition monitoring using a thermal conductivity sensor*. in *Proc. GMe Forum (Vienna, USA)*. 2005.
- [5] Bošković, P., et al., *Detection of lubrication starved bearings in electrical motors by means of vibration analysis*. Tribology international, 2010. 43(9): p. 1683-1692.
- [6] Mokšin, V., A. Kilikevičius, and I. Tetsman, *7. Investigation of vibrational behavior of friction pair under starved lubrication conditions*. 2013.

- [7] Jamšek, J., Đ. Juričić, and P. Boškosi, *Lubrication starved bearings detection in electrical motors vibration signals by means of wavelet bispectral analysis*.
- [8] de Almeida, R.G. and L.R. Padovese, *Characterization of oil viscosity alterations in a gearbox through vibration signal analysis*. in *17th International Congress of Mechanical Engineering (COBEM2003)*. 2003. November 10-14, Sao Paulo, SP.
- [9] Abusaad, S.B., Ahmed Brethee, Khaldoun F Gu, Fengshou Ball, Andrew, *Investigating the Effect of Water Contamination on Gearbox Lubrication based upon Motor Control Data from a Sensorless Drive*. In: *Proceedings of the 3rd international workshop and congress on eMaintenance: June 17-18 Luleå, Sweden : eMaintenance, Luleå tekniska universite, Luleå*, pp. 61-67,, 2014.
- [10] Tian, X.G., Fengshou Rehab, Ibrahim Abdalla, Gaballa M Ball, Andrew D, *An MSB based robust detector for bearing condition monitoring*, in *28th International Congress of Condition Monitoring and Diagnostic Engineering Management (COMADEM 2015)*. 2015: Buenos Aires, Argentina.
- [11] Rehab, I.T., Xiang Gu, Fengshou Ball, Andrew, *The fault detection and severity diagnosis of rolling element bearings using modulation signal bispectrum*, in *In: Eleventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*. 2014: 10th-12th June 2014, Manchester, UK.
- [12] G. W. Stachowiak, A.W.B., *Engineering tribology*, ed. F. Edition. 2005: Cambridge University Press.
- [13] Peng, Z., N.J. Kessissoglou, and M. Cox, *A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques*. *Wear*, 2005. **258**(11): p. 1651-1662.
- [14] Antoniadou, I., et al., *A time-frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions*. *Mechanical Systems and Signal Processing*, 2015. **64**: p. 188-216.
- [15] Jakoby, B. and M.J. Vellekoop, *Physical sensors for water-in-oil emulsions*. *Sensors and Actuators A: Physical*, 2004. **110**(1): p. 28-32.
- [16] Johnsen, E.E. and H.P. Rønningsen, *Viscosity of 'live' water-in-crude-oil emulsions: experimental work and validation of correlations*. *Journal of Petroleum Science and Engineering*, 2003. **38**(1): p. 23-36.
- [17] Liu, G.P., Robert G, *Impact of tooth friction and its bending effect on gear dynamics*. *Journal of Sound and Vibration*, 2009. **320**(4): p. 1039-1063.
- [18] Brethee, K.F.G., Fengshou Ball, Andrew D, *Frictional effects on the dynamic responses of gear systems and the diagnostics of tooth breakages*. *Systems Science & Control Engineering*, 2016. **4**(1): p. 270-284.
- [19] Hamel, M.A., *Condition Monitoring of Helical Gears Using Acoustic Emission (AE) Technology*. 2013, PhD Thesis, Cranfield University.
- [20] Xu, H., *Development of a generalized mechanical efficiency prediction methodology for gear pairs*. 2005, The Ohio State University.

Weak Fault Diagnosis of Bearings Using a Cyclic Deconvolution Method by Combining SMHD with MEDA

Yonghao Miao¹, Ming Zhao¹, Jing Lin², Jun Zhou³, Jianming Wang³

¹ Shaanxi Key Laboratory of Mechanical Product Quality Assurance and Diagnostics, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

² State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, China

³ Qinghai Huading Heavy Machine Tool CO., LTD, Xining 810000, China

ABSTRACT

This paper proposes a cyclic deconvolution method by combining sparse maximum harmonics-noise-ratio deconvolution (SMHD) with minimum entropy deconvolution adjusted (MEDA) for the weak fault diagnosis of rolling element bearings. In the method, the main purpose of using two deconvolution processes to counteract the influence of the transmission path. SMHD is regarded as the pretreatment to enhance the periodic nature of the bearing fault signals. Subsequently, an improved MED, MEDA which overcomes the disadvantage to deconvolve the spurious impulse from the filtered signal, is used to clarify the periodic impulse components of the bearing signals. The proposed method not only overcomes the limitations associated with traditional deconvolution methods, MED and maximum correlated kurtosis deconvolution (MCKD), but has a better visual inspection, even if the fault period is not provided in advance. Moreover, the efficiency of the proposed method is verified by simulation and bearing experimental data. The results show that the proposed method could effectively suppress the heavy background noise and extract the weak fault characteristics from the vibration signals compared with MED and MCKD.

Keywords: rolling element bearing, weak fault diagnosis, deconvolution

Corresponding author: Jing Lin (email: jinglin@mail.xjtu.edu.cn)

1. INTRODUCTION

The rolling element bearings are widely used as key components in various rotating machinery systems. A tiny defect in a bearing will develop into a considerable security risk if not being detected timely. The failure of rolling element bearings will cause unscheduled downtime or even devastating disaster during the high-speed operation. Therefore, detecting and diagnosing bearing incipient faults are very important in machinery condition monitoring and diagnostics [1]. Since vibration signals collected from fault bearings carry rich information on machine health conditions, vibration measurements are widely applied to monitor the bearing condition [2].

A common problem encountered in the fault detection and diagnosis of bearings is that vibration signals from the monitoring sensors are often contaminated by background noise or interference [3]. To enhance and extract the defect features, several advanced time-frequency analysis techniques, such as Short Time Fourier Transform (STFT) [4], wavelet transform (WT) [5], and Empirical Mode Decomposition (EMD) [6] have been introduced and applied to fault diagnosis of rotating machinery. But it should be noted that these transformations are much more complicated compared with the deconvolution methods. Since the deconvolution methods have the simpler conception and possess a significant advantage in counteracting the influence of the transmission path and enhancing the fault impulses. For instance, the minimum entropy deconvolution (MED) method is designed to reduce the spread of impulse response functions (IRFs), to obtain signals closer to the original impulses that gave rise to them [7]. Endo *et al.* [8] first introduced and successfully applied MED in the field of fault detection. Subsequently, the method was applied by Sawalhi *et al.* [9] to extract the bearing fault combined with spectral kurtosis (SK) and the technique showed good

performance. However, MED may fail to extract the desired ones when it encounters the large random impulses especially for the random impulse and the impulse train that resided in different resonance frequencies. For this reason, an improved deconvolution technique, maximum correlated kurtosis deconvolution (MCKD), which takes advantage of the periodicity of the faults, was proposed to overcome the limitation. It is obvious that MCKD [10] shows more significant performance advantage in extracting weak features from noisy signals when the precise fault period is provided in advance. Therefore, the technique is considerably appropriate to be a pretreatment signal analysis procedure. Jia *et al.* [11] takes advantage of MCKD as a pretreatment to enhance the periodic nature of the bearing fault signals leading to the result that SK can select the more accurate resonant frequency band of the signal filtered by MCKD. Considering the deconvolution process of MCKD containing a resample operation, which will change the sample frequency of the filtered signal, few literature uses the envelop analysis combining with MCKD directly. In addition, its ability to detect the bearings faults can be further enhanced in the situation where the noise and interference are highly complicated.

To address the above issues, this paper proposes a cyclic deconvolution method by using two deconvolution processes. The cyclic deconvolution method introduces a new pretreatment, sparse maximum harmonics-noise-ratio deconvolution (SMHD), and applies MEDA as the after treatment. Inspired by MED and extended to the technique objective function method (OFM) [12], SMHD employs the new index, harmonics-to-noise ratio (HNR), as the objective function by iteratively selecting an FIR filter to maximum the HNR of the filtered signal. To further enhance the periodic components, a spare factor is designed to suppress the noise and improve the signal-to-noise ratio (SNR) of the filtered signal after each filter. The new method, CDM, takes full advantage of the merits of SMHD and MEDA to reach the goal of the accurate diagnosis of the weak faults. First, SMHD is applied to highlight the periodic components of the vibration signals. Then MEDA is used to accurately extract the fault impulses from the filtered signal. Compared with the conditional deconvolution methods, CDM is expected to be more effective.

The rest of this paper is organized as follows. In section 2, the review of the deconvolution methods is given, and the limitations of the methods are discussed. The new method CDM is proposed in section 3. Section 4 presents the procedure and validation of CDM using simulated bearing data. In section 5, the real dataset from the Case Western Reserve University (CWRU) Bearing Data Center further highlights the advantages of CDM. Finally, the conclusion is given in section 6.

2. DECONVOLUTION METHODS

When a localized fault occurs in rolling element bearing, the fault impulse will excite resonances at much higher frequencies than the vibration generated by other machinery components, and this resonant response is considered to be amplitude modulated at the characteristic defect frequency. Vibration signals of faulty bearings are usually composed of impulses, fault-excited resonance, cyclic interferences from other mechanical or/and electrical components, and background noise. The fault-induced impulses are often masked by these interferences and noise, and therefore the envelope of impulses cannot be directly extracted by some conventional methods. Furthermore, an excessively long and complex transmission path between bearings and transducers would affect the frequency-response function of the vibration signal. It leads to the result that the amplitude of impulse would be attenuated and the time would be prolonged. As a result, the impulse signals are submerged into the severe background noise easily [13].

Deconvolution through the iterative process aims at approximately inverting this transmission process to improve the time resolution in order to restore impulse signals generated by faults. Figure 1 illustrates the deconvolution process. The vibration signal x from the transducer is comprised of the fault impulses of bearings in original form g and the noise interference n which need to go through the respective transmission paths. h_g and h_n represent the FIR filter approximations to their respective transfer functions under stable assumption. Therefore, the time domain approximation forms in terms of convolution of the vibration signal x can be illustrated as:

$$x = g * h_g + n * h_n \quad (1)$$

To restore impulse signals, OFM is applied in the deconvolution methods (MED and MCKD) whose basic idea is to find an inverse filter that counteracts the effect of the transmission path, by assuming that the original excitation was approximatively impulsive and periodic, and thus having high kurtosis or correlated kurtosis (CK). The filtered signal by optimizing the inverted filter coefficients f_k to make the objective function reach the maximum value and the result y is expected to be the fault signal g .

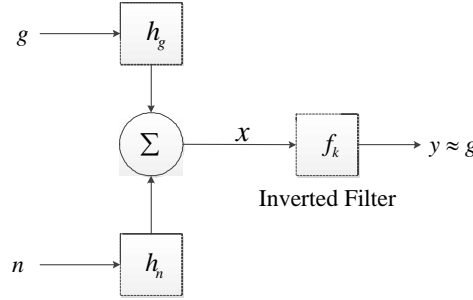


Figure 1. The deconvolution process.

Wiggins [12] introduced the OFM and firstly derived the mathematics of one-dimensional MED. The derivation process is given in the following description. The signal's kurtosis is designed to be the OF:

$$K = \frac{\frac{1}{N} \sum_{n=1}^N (y_n - \mu_y)^4}{\left\{ \frac{1}{N} \sum_{n=1}^N (y_n - \mu_y)^2 \right\}^2} - 3 \quad (2)$$

where y_n is the output signal and it is assumed to be zero-mean, i.e., $\mu_y=0$; N is the sampling number. Therefore, the following expression can be given:

$$\max_f K = \max_f \frac{\sum_{n=1}^N y_n^4}{\left(\sum_{n=1}^N y_n^2 \right)^2} \quad (3)$$

$$\frac{\partial y(n)}{\partial f(l)} = x(n-l) \quad (4)$$

where $f(l)$ is the filter coefficient, $l=1,2,\dots,L$; L is the length of the filter; $x(n)$ is the input signal.

By combining with the inverse filtering expression (4) and taking the derivatives of (3) with respect to filter coefficients f and solving it equal to zero, the final iterative expression can be derived. The detailed step of this procedure is described in [8]. Compared with MED, MCKD which aims to deconvolve periodic impulse faults from a machine vibration signal takes the periodicity of the fault into consideration. Different from MED, MCKD introduced a new norm, CK, and was designed to use CK as the OF to iteratively choose the optimum filter coefficients to maximize CK of the filtered signal. CK of M - shift was defined [10]:

$$CK_M(T_s) = \frac{\sum_{n=1}^N \left(\prod_{m=0}^M y_{n-mT_s} \right)^2}{\left(\sum_{n=1}^N y_n^2 \right)^{M+1}} \quad (5)$$

where $y_n = \sum_{k=1}^L f_k x_{n-k+1}$, $x_n=0$ and $y_n=0$ for $n \neq 1, 2, \dots, N$. f_k , L and N denote the filter coefficients, the length of the filter and the sampling number of the input signal, respectively. T_s is the corresponding sampling number of fault period provided in advance. Similarly, MCKD also applies the OFM. By combining with the inverse filtering expression (4) and taking the derivatives of (6) with respect to filter coefficients f and solving it equal to zero, the final iterative expression can be used to obtain the optimum filter coefficients.

$$\max_f CK_M(T_s) = \max_f \frac{\sum_{n=1}^N \left(\prod_{m=0}^M y_{n-mT_s} \right)^2}{\left(\sum_{n=1}^N y_n^2 \right)^{M+1}} \quad (6)$$

Ref. [10] describes in detail the iterative process.

There is no doubt that MCKD shows the more obvious performance advantage than MED. Furthermore, MCKD has overcome some shortcomings of MED. For instance, MCKD can be used to process the signal experienced the band-pass filter and the signal contains large random impulses, while MED is sensitive to them. However, the limitation that MCKD depends strongly on the precise estimation of the fault period in advance has restricted its applications. To overcome the disadvantages of the conditional deconvolution method, MED and MCKD, Miao *et al.* [14] proposed SMHD by introducing a novel index, HNR, which is particularly effective for evaluating the degree of periodicity of a signal and applied successfully in the field of bearing fault diagnosis, voice activity detection and automatic speech recognition, etc. [15, 16]. HNR is designed to estimate the period and to be the OF in the OFM. By maximizing HNR, the periodic component is divided from the mixed signal. In addition, combined with the sparse process, SMHD has the ability to further highlight the periodic components and restrain the others. Therefore, SMHD can be considered as a feature enhancing technique without requiring any prior knowledge [14].

3. CYCLIC DECONVOLUTION METHOD

The conventional deconvolution methods (MED and MCKD) have been applied to machinery fault diagnosis for years and has achieved good results [8-10]. In some cases, however, MED does not attain the expected goal of machine diagnosis. It is obvious MED is sensitive to the signal experienced the band-pass filtered signal and the large random impulse in figure 2. It can be seen that the filtered signal of the one experienced the band-pass filter by MED in (a), and (b) illustrates the filtered signal of the one contains a large random impulse by MED. To overcome these problems, an improved MED, MEDA, is proposed by choosing the valid part of the signal to deconvolve to avoid the impulse near the start of the signal [17]. SMHD, which has overcome the trend that the enhanced components only contain the large random aperiodic impulses rather than the periodic impulses we are interested in, is developed by considering the periodic nature of the fault impulses. Nevertheless, in some weak faults using SMHD to detect the impulses of the signals is also difficult since potential periodic impulses are often overwhelmed by unexpected heavy noise and interferences. SMHD needs to choose the appropriate sparse threshold value to enhance the periodic impulses. Therefore, to extract weak fault more robustly, SMHD is considered as the pretreatment method, and MEDA is regarded as the aftertreatment in the new method. It is can be ensured that CDM can take full advantage of the merits of SMHD and MEDA to reach the goal of the accurate diagnosis of the weak faults.

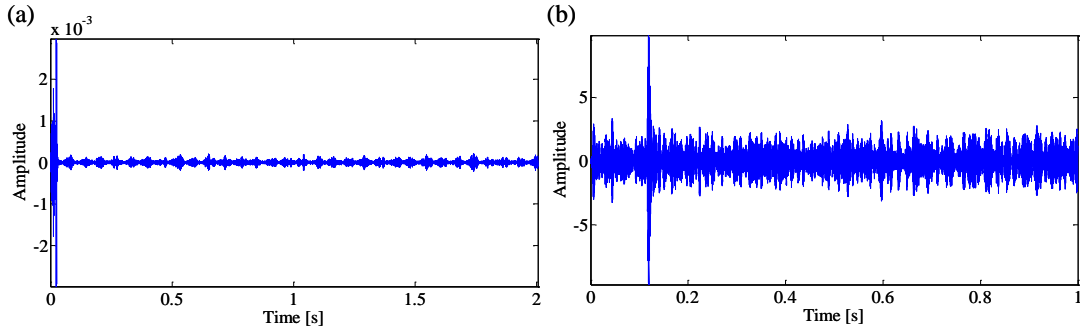


Figure 2. The filtered signal of (a) the one experienced the band-pass filter and (b) the one contains a large random impulse by MED.

Based on the analysis above, this paper proposes a cyclic deconvolution method for weak fault diagnosis of rolling element bearings. The details of CDM are described as follows.

Step 1: The original vibration signal x is input. x_1 is obtained by SMHD filter.

Step 2: The signal x_1 experiences MEDA filter and the filtered signal x_2 is output.

Step 3: Choose the optimal filtered signal x_2 and input it to the next step.

Step 4: Detect the fault characteristic frequency in the envelope spectrum and diagnose the fault types.

4. SIMULATION

To highlight the advantages of the new method, simulated signals are generated according to the vibration model developed by McFadden[18]. The mathematical model of defective rolling element bearing can be expressed as:

$$x(t) = \underbrace{\sum_{i=1}^M A_i s_i(t - iT - \tau_i)}_I + \underbrace{\sum_{j=1}^{M_1} D_j s_j(t - jT)}_{II} + \underbrace{\sum_{k=1}^{M_2} P_k \cos(2\pi f_k t + \beta_k)}_{III} + \underbrace{n(t)}_{IV} \quad (7)$$

The simulated signal is composed of four terms. The first term is the periodic impulses caused by faults. The ball pass frequency of outer race (BPFO) and the natural frequency of the bearing system are set 29 Hz and 2000 Hz, respectively. τ_i is the random fluctuation because of the random slip in bearings round fault period T and its variance is set 0.01. The second term represents vibration components from gearbox and other related rotating parts. The two harmonics of shaft in the signal, and their amplitudes, phases and instantaneous rotating frequency of shaft are $P_1 = 0.005$, $P_2 = 0.005$, $\beta_1 = \pi/6$, $\beta_2 = -\pi/3$, $f_1 = 5$, $f_2 = 10$, respectively. The third term represents the aperiodic impulse from extraneous sources. The resonance frequency of the aperiodic impulses is set 2645 Hz. The last term is the -15 dB Gaussian noise with zero-mean. In addition, the data length and the sampling rate are 1 s and 20k Hz, respectively.

It is obvious that the periodic impulses are masked by noise from figure 3(a). The envelope spectrum can discover the second harmonic of BPFO, while the signal-to-noise ratio (SNR) in figure 3(b) is much low. Without loss of generality, in this paper the length of the filter and the maximal iteration count are same in the all deconvolution methods and they are 100 and 30, respectively. The third-shift MCKD method is the recommended setting in [10]. To highlight the superiority, sparse threshold value is set RMS of the original signal in CDM in all simulations and experimental data. Compared with the results of the two methods (MED and MCKD), CDM is more suitable and effective to weak characteristics extraction from noisy signals. Since CDM can discover the fault characteristics from a noisy signal with a significant SNR, MCKD fails from figures 4 and 5.

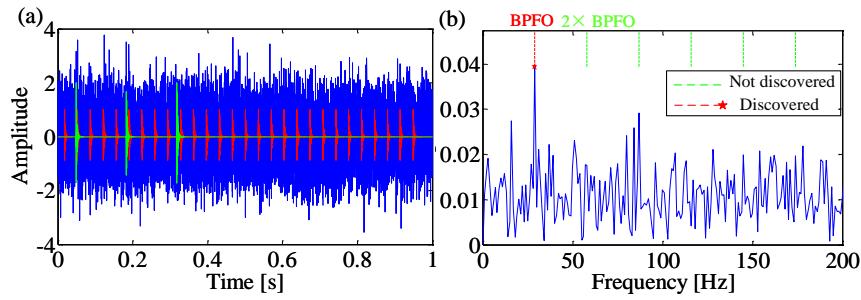


Figure 3 (a) The Mixed signal (red line represents pure periodic impulse signal; green line represents aperiodic impulses); (b) The envelope spectrum of (a).

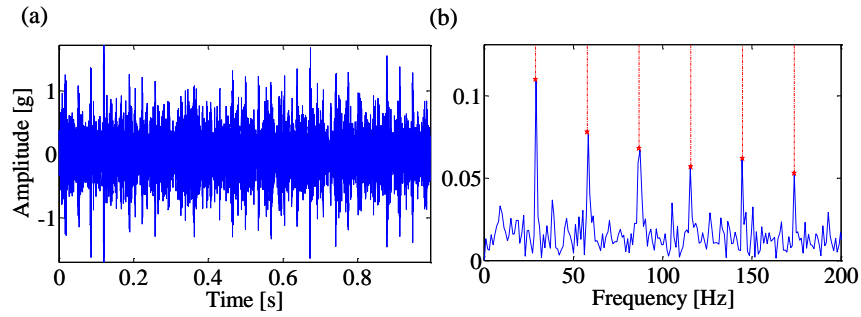


Figure 4. (a) Filtered signal using CDM, and (b) the envelope spectrum of (a).

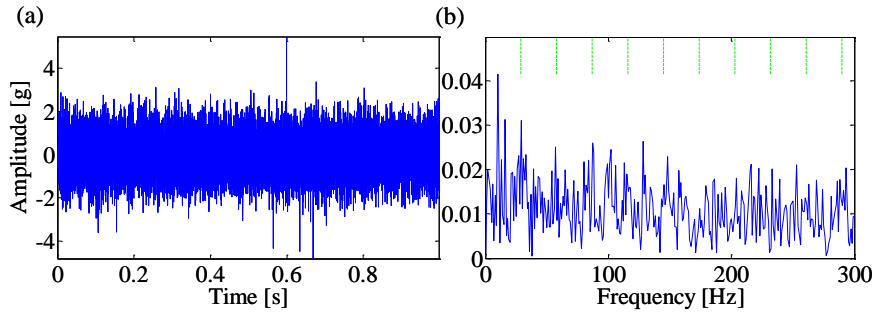


Figure 5. (a) Filtered signal using MCKD, and (b) the envelope spectrum of (a).

5. EXPERIMENT

Over the last decade, the bearing data [19] from the CWRU Bearing Data Center has been validated in many researches, which appears in several academic journals. Recently, Wade A. Smith *et al.* [20] made a very thorough study on the data from CWRU Bearing Data Center and categorized it appropriately. In the article, three established diagnostic techniques which include envelope analysis, cepstrum prewhitening and benchmark method introduced in [7] were applied to the entire CWRU datasets, and the diagnostic outcomes were provided and discussed in detail. A dataset 199DE was considered to be partially diagnosable by envelope analysis. To further verify the broader applicability and robustness of CDM, this paper will use the dataset and try to discover the fault characteristics.

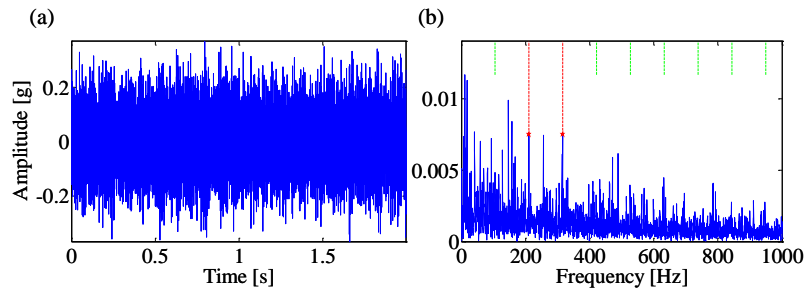


Figure 6. (a) Raw signal and (b) envelope spectrum analysis in experiment.

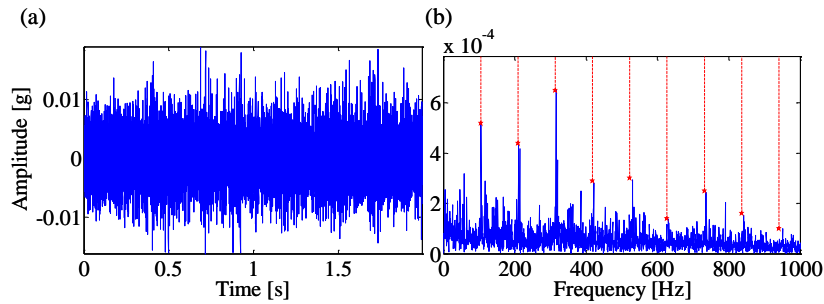


Figure 7. Results of applying CDM to experimental signal: (a) filtered signal; and (b) envelope spectrum.

The length of the signal used for analysis is 2 s. The rotating speed provided is 1749 RPM and the calculated BPFO is 104.4969 Hz. Figure 6 gives the raw signal and the envelope spectrum analysis. The periodic impulses are completely buried by noise and its envelope spectrum is masked heavily though the harmonics of BPFO can be faintly discovered. For comparison, CDM and MCKD are applied to the same outer race fault signal. The filtered signals and the envelope spectrum of the signals show fault related signatures in figures 7 and 8. From the results, the filtered signals and envelope spectra obtained by CDM have better visual inspection ability than MCKD. As a matter of fact, MCKD fails to discover the fault characteristics.

However, it is interesting to find that the all peak values in the envelope spectrum cannot correspond to the calculated BPFOs accurately in figure 7(b). Therefore, the peak values in the envelope spectrum are regarded as the fault characteristic and its harmonics and applied in CDM and MCKD again. Figures 9 and 10 show the results after changing the fault frequency from 104.4969 Hz to 105.5 Hz. It is obvious that 105.5 Hz is the real fault frequency from the results. This phenomenon provides convincing evidence that the provided rotating speed is imprecise. In this situation, CDM still has more significant performance advantage than MCKD. It is obvious seen that the fault impulses are considerably clear from the filtered signal by CDM in figure 9(c). However, the envelope spectrum of the original vibration signal is difficult to discover BPFO, it can be found that the impulses caused by the fault do exist but are very weak.

Through the comparisons between CDM and MCKD in the experiment, it is appropriate to draw the following some conclusions. There is no doubt that CDM is more suitable and effective to weak characteristics extraction from noisy signal. In addition, MCKD extremely depends on the precise estimation of the fault period in advance restricts its application range. Whereas the experiment highlights the another merit, the results are not affected by the imprecise prior knowledge, of CDM.

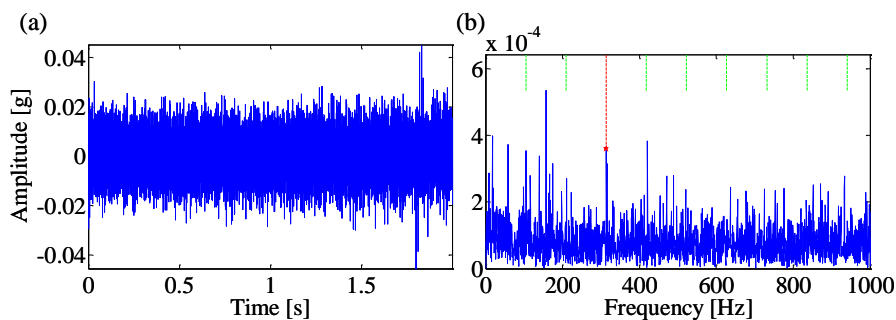


Figure 8. Results of applying MCKD to experimental signal: (a) filtered signal; and (b) envelope spectrum.

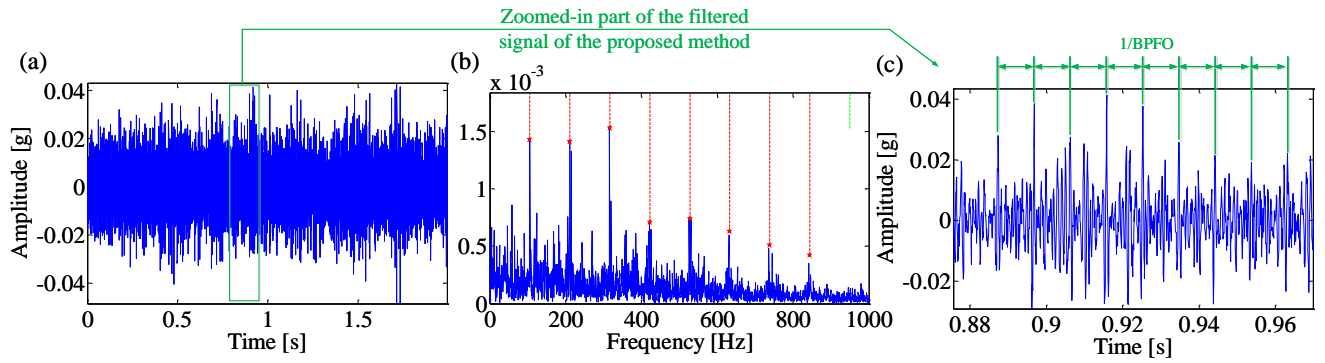


Figure 9. Results of applying CDM to experimental signal after changing the period: (a) filtered signal; (b) envelope spectrum; and (c) the zoomed-in part of (a).

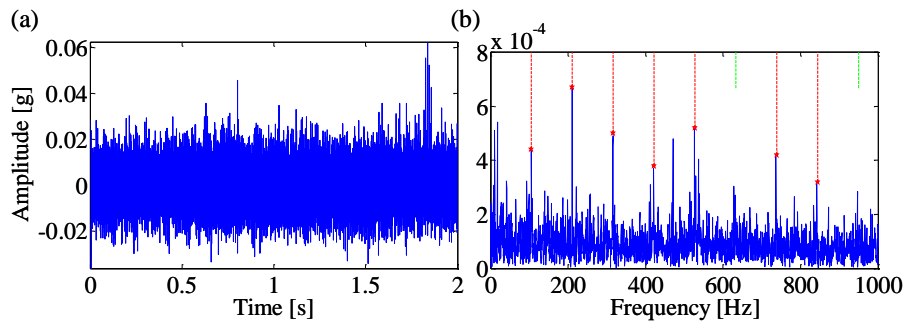


Figure 10. Results of applying MCKD to experimental signal after changing the period: (a) filtered signal; (b) envelope spectrum.

6. CONCLUSION

Enhancement and extraction of the weak signature from the noisy signal are crucial to condition monitoring and fault prognostics. However, the fault features are often very weak and masked by the background noise. The deconvolution methods have a significant performance advantage in counteracting the effect of the transmission path and enhancing the fault impulses. This paper proposes a novel cyclic deconvolution which aims to extract the weak fault signature that is difficult for conventional deconvolution methods. The new method takes full advantage of the merits of SMHD and MEDA. That is, SMHD has a significant performance advantage as the pretreatment gain from enhancing the periodic nature of the bearing fault signals and MEDA is appropriate to clarify the periodic impulse components of the bearing signals. In the cases of simulated signal, compared with MCKD and MED, CDM has a better detection rate for fault frequency even though there is much interference. Moreover, the efficiency of CDM is verified by bearing data from CWRU Bearing Data Center.

ACKNOWLEDGEMENT

This project is supported by the National Science and Technology Major Project (2014ZX04001051), which is highly appreciated by the authors.

REFERENCES

- [1] J. Wang, Q. He, F. Kong, Adaptive multiscale noise tuning stochastic resonance for health diagnosis of rolling element bearings, *Instrumentation and Measurement, IEEE Transactions on*, 64 (2015) 564-577.
- [2] R.B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*, John Wiley & Sons, 2011.
- [3] C. Wang, R.X. Gao, Wavelet transform with spectral post-processing for enhanced feature extraction [machine condition monitoring], *Instrumentation and Measurement, IEEE Transactions on*, 52 (2003) 1296-1301.

- [4] L. Satish, Short-time Fourier and wavelet transforms for fault detection in power transformers during impulse tests, *Science, Measurement and Technology, IEE Proceedings-, IET*, 1998, pp. 77-84.
- [5] Z. Peng, F. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, *Mechanical systems and signal processing*, 18 (2004) 199-221.
- [6] Y. Lei, J. Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, *Mechanical Systems and Signal Processing*, 35 (2013) 108-126.
- [7] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—A tutorial, *Mechanical Systems and Signal Processing*, 25 (2011) 485-520.
- [8] H. Endo, R.B. Randall, Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter, *Mechanical Systems and Signal Processing*, 21 (2007) 906-919.
- [9] N. Sawalhi, R.B. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, *Mechanical Systems and Signal Processing*, 21 (2007) 2616-2633.
- [10] G.L. McDonald, Q. Zhao, M.J. Zuo, Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection, *Mechanical Systems and Signal Processing*, 33 (2012) 237-255.
- [11] F. Jia, Y. Lei, H. Shan, J. Lin, Early Fault Diagnosis of Bearings Using an Improved Spectral Kurtosis by Maximum Correlated Kurtosis Deconvolution, *Sensors*, 15 (2015) 29363-29377.
- [12] R.A. Wiggins, Minimum entropy deconvolution, *Geophysical Research Letters*, 16 (1978) 21-35.
- [13] Y. Miao, M. Zhao, J. Lin, Y. Lei, Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings, *Mechanical Systems and Signal Processing*, 92 (2017) 173-195.
- [14] Y. Miao, M. Zhao, J. Lin, X. Xu, Sparse maximum harmonics-to-noise-ratio deconvolution for weak fault signature detection in bearings, *Measurement Science and Technology*, 27 (2016) 105004.
- [15] P. Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, *Proceedings of the institute of phonetic sciences, Amsterdam*, 1993, pp. 97-110.
- [16] X. Xu, M. Zhao, J. Lin, Y. Lei, Periodicity-based kurtogram for random impulse resistance, *Measurement Science and Technology*, 26 (2015) 085011.
- [17] G.L. McDonald, Q. Zhao, Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to vibration fault detection, *Mechanical Systems and Signal Processing*, (2016).
- [18] P. McFadden, J. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, *Journal of sound and vibration*, 96 (1984) 69-82.
- [19] K.A. Loparo, Bearings vibration data set, Case Western Reserve University Bearing Data Center Website, OH, <http://csegroups.case.edu/bearingdatacenter/home>, 2003.
- [20] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, *Mechanical Systems and Signal Processing*, 64-65 (2015) 100-131.

Quantitative Detection of the Steel Surface Cracks Based on Eddy Current Pulsed Thermography

Jiao-jiao Ma¹, Xiaoyu Xu¹, Dong Zhen¹, Hao Zhang¹, Zhan-qun Shi^{1,*}

¹School of Mechanical Engineering, Hebei University of Technology, Tianjin, China.

Corresponding author: z_shi@hebut.edu.cn

ABSTRACT

Eddy Current Pulsed Thermography(ECPT) is a developed technique in Non-Destructive Testing and Evaluation (NDT&E). It is widely used in detecting metal surface and sub-surface defects. In the present researches, many thermal feature extraction methods have been used to achieve more valuable hidden information. At the same time, a quantitative analysis strategy is proposed for determining the size of the deflection based on Eddy Current Pulsed Thermography. In this paper, the Q235 was selected as the analysis materials, with multiple surface defects detection, was studied by using eddy current pulsed thermography. Based on the theoretical and experimental studies, the results show that the pulsed eddy current thermal imaging technology can detect the different position defects both on the qualitative and the quantitative evaluation effectively. It can serve as guidelines to improve the visibility and detectability of defects with different defects in the future.

Keywords: Eddy Current Pulsed Thermography , Q235, Surface Cracks, Quantitative Evaluation

Corresponding author: Prof. Z. Shi (z_shi@hebut.edu.cn)

1. INTRODUCTION

Eddy Current Pulsed Thermography(ECPT) as a developed technique in Non-Destructive Testing and Evaluation (NDT&E) has advantages of both non-destructive and visual detection. This technique directly makes the conductive material generate heat by the excitation method of induction heating. Meanwhile, the thermal infrared imager records the surface temperature of the specimen. Due to the influence of the surface defects on the distribution of the eddy current, the characteristics of the cracks in the infrared image are obvious. Also, the important information will be obtained by the thermal image sequences such as the spatial, time and frequency domains, thus it is very suitable for the detection of metal surface cracks [1,2].

Currently, many thermal feature extraction methods have been used to achieve more valuable hidden information. At the same time, the research on eddy current thermography has developed from qualitative analysis to quantitative analysis. The ability of ECPT to detect small cracks in components of complex geometry was investigated in [3]. The variation law of temperature near the crack of ferromagnetic material under low frequency induction heating was studied in [4]. Genetic algorithm was used to detect the radius and depth of the defects on the lower surface of the aluminum plate under COMSOL simulation [5]. Frequency optimization of excitation current for notch detection on aluminum material was studied in [6]. The relationship between the temperature of near the surface crack and the crack depth and width in ANSYS simulation was studied in [7]. However, most of the research are depended on a single defect on the surface of the metal. The quantitative assessment and error analysis about multiple defects on the metal surface remain as challenges. In this paper, the Q235 was chosen as the case study materials. The test samples with three different positions crack defects on the surface is used in the experimental study. Based on the temperature image collected in the experiments, the boundary position and the relative position of the defects can be extracted by using a novel feature extraction method.

The remainder of this paper is organized as follows. Section 2 presents the experiment set-up and samples used for demonstration. Section 3 shows the experiment results, and the relative positions of several defects are extracted based on the temperature difference method. In the end, the conclusion is drawn in Section 4.

2. EXPERIMENTAL SET-UP

The experimental system of defects detection using pulsed eddy current thermography is illustrated in figure 1. The system includes an infrared thermal camera, a computer, an induction heating unit with an exciting coil. In the test of the sample, the infrared thermal imager is used to record the changing information of the heat distribution on the specimen surface. However, due to the lack of experimental conditions that the infrared thermal imager is relatively low in the two aspects of pixel and frequency, as listed in table 1, the size of the specimen and the defect has special requirements. Beyond that the positions of three groove shape cracks are made on the surface of the specimen as shown in figure 2(a) and the actual specimen is shown in figure 2(b).

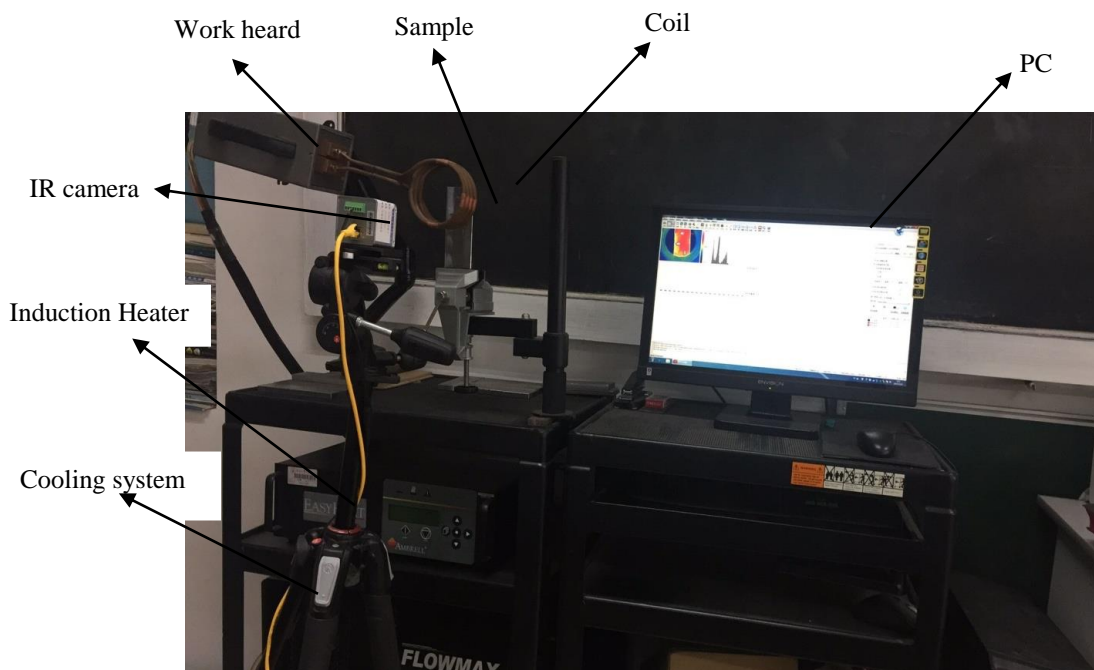


Figure 1. The experiment system

Table 1: The size of the sample and defects

| Size | The carbon steel(Q235) | Crack defects |
|-------------------|------------------------|---------------|
| Length/mm | 150 | 10 |
| Width/mm | 50 | 5 |
| Height(Depth) /mm | 10 | 2 |

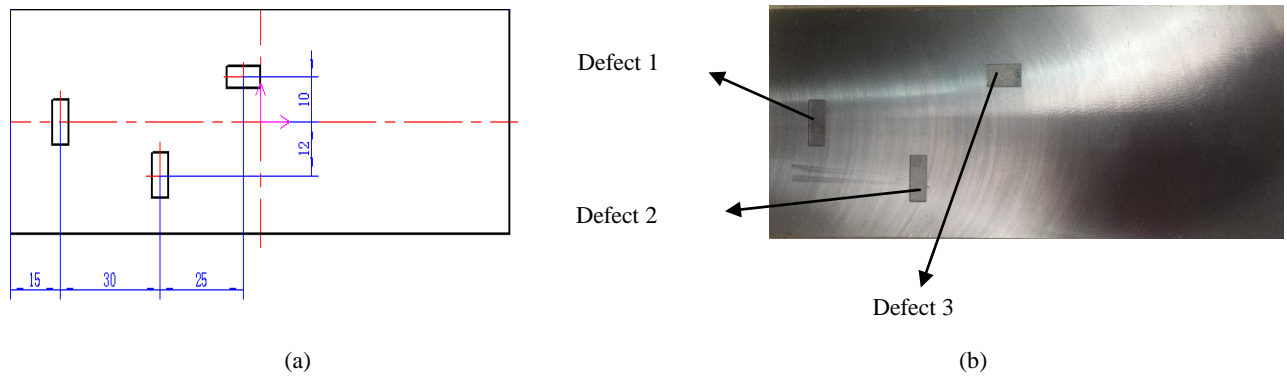


Figure 2. The diagram of specimen (a) and the carbon steel with multiple defects (b)

In particular, the infrared range of the thermal imager is from 7.5 microns to 14 microns, the infrared pixel is $80 * 60$, and the frame rate is about 50 Hz. Besides, the induction heating system is Cheltenham Induction Heating Easy Heat 0224 with an output power of 1.2~2.4 kW across the frequency range of 150~400 kHz. The current strength and excitation time can be set by the detector using the control panel of the induction heating unit. The experimental parameters are listed in table 2. In the experiments, the relative position between the exciting coil and the specimen is shown in figure 3.

Table 2: The Excitation Parameters

| Excitation parameters | Numerical |
|--------------------------|-----------|
| Current strength/A | 260 |
| Excitation frequency/kHz | 380 |
| Life-off/mm | 20 |
| Excitation time/s | 2 |



Figure 3. Relative position of the specimen and the exciting coil

3. RESULTS AND DISCUSSION

According to the principle of electromagnetic induction, the conductive structures detected will generate eddy currents when the exciting coil closes to the conductive structures detected. While, surface defects affect the distribution of eddy current field, and then influence the distribution of surface temperature field. With the gradual heating of the excitation time, the eddy current gathers in the defect part of the specimen and the temperature of the defect part increases gradually, leading to the high-temperature effect. Therefore, some bright area appears on the surface of the specimen, namely defect. In the whole heating process, infrared thermal imager records the variation of temperature on the surface of the specimen. As shown in figure 4, where, figure 4(a) is the initial temperature of the image, including coil temperature, specimen temperature and ambient temperature. the different distribution of surface temperature field under the excitation time with 0.5s interval are shown in figure 4(b) to 4(f). With the increase of heating time, it is clear that the temperature of the specimen is in sharp contrasted with the ambient temperature and the coil temperature. After heating 2s, the temperature data matrix is analyzed in horizontal and vertical direction, and the results are shown in figure 5 and figure 6.

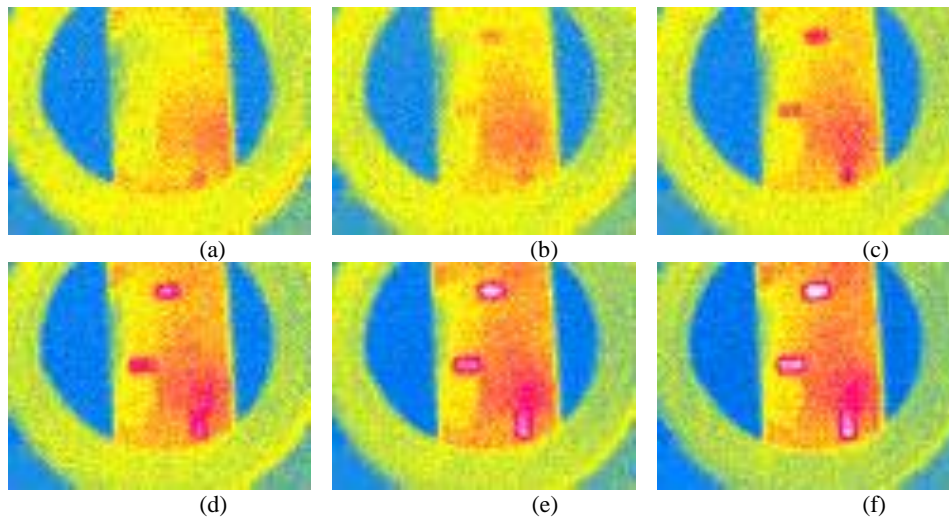


Figure 4 Infrared thermography under the different heating time

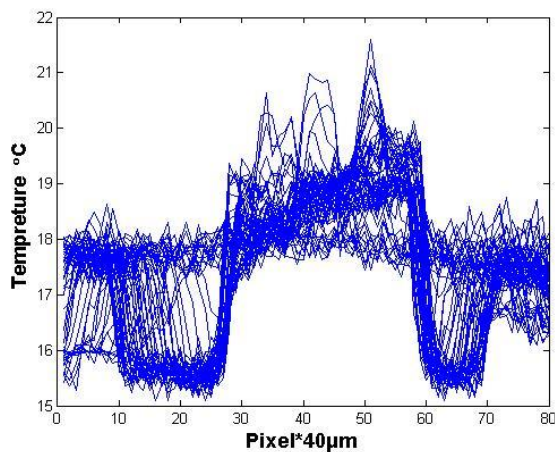


Figure 5. Horizontal temperature of infrared thermography

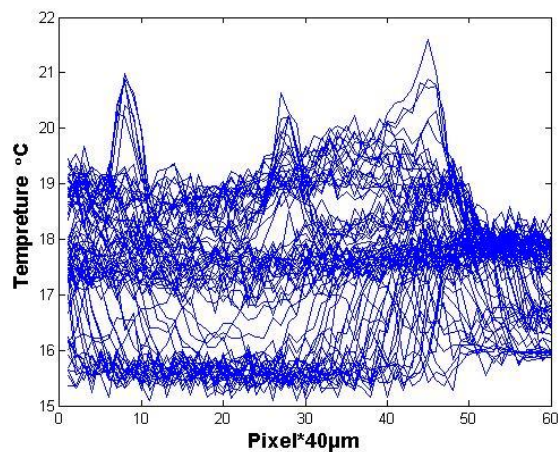


Figure 6. Vertical temperature of infrared thermography

In order to realize the quantitative analysis of the defect on the specimen, the temperature difference method is used to process the initial temperature data and the result is illustrated in figure 7 and figure 8. The temperature difference method is the difference of adjacent temperature, so the difference of environment temperature is small and almost is zero. While, the temperature difference between the heated specimen and the ambient temperature is high. According to the location of the specimen and coil in the imaging and the temperature data obtained by the infrared thermal imager, it is obvious to show the boundary position of the specimen with the tip of temperature difference. As shown in figure 7 and figure 8, the horizontal boundary is located between 23rd pixel and 61st pixel and the vertical boundary is located

between 5th pixel and 46th pixel. Extracting effective temperature data is good for reducing the workload of data processing, thus, the next work is to analyse this new temperature data matrix.

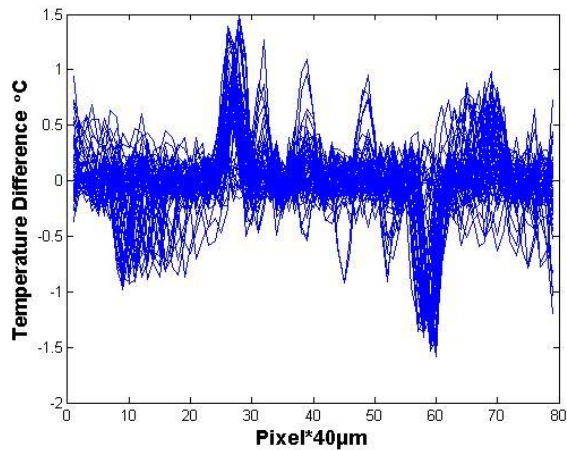


Figure 7. Horizontal temperature of infrared thermography

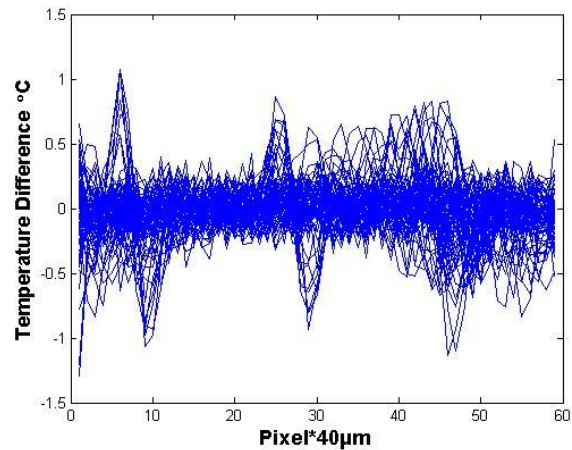


Figure 8. Vertical temperature of infrared thermography

To detect the size of the defect and the relative position of the three defects accurately, it is necessary to eliminate the influence of the coil temperature and the ambient temperature from the initial data. Then, the temperature data matrix is extracted from three rows of the highest temperature at the place, which these defects are located on, and also the above new temperature data matrix is considered. The temperature curves of the three defects are found in figure 9 and figure 10. After using the temperature difference method, the results are shown in figure 11 and figure 12.

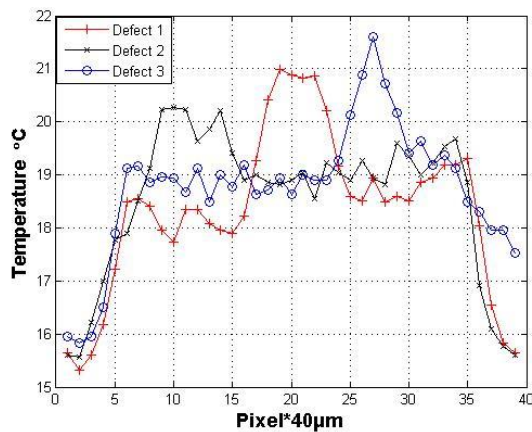


Figure 9. Horizontal temperature of different defects

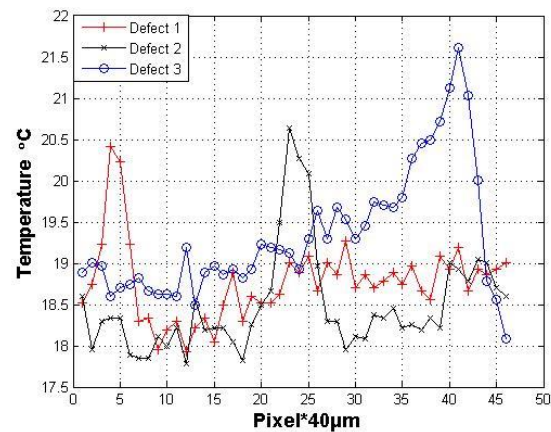


Figure 10. Vertical temperature of different defects

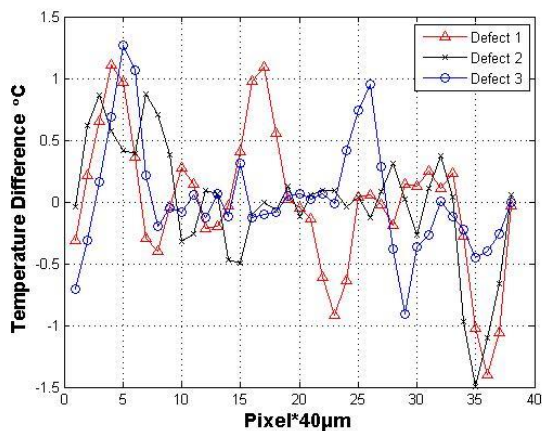


Figure 11 Horizontal temperature difference of defects

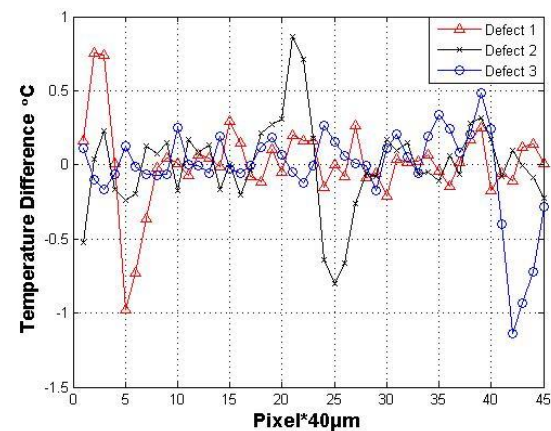


Figure 12 Vertical temperature difference of defects

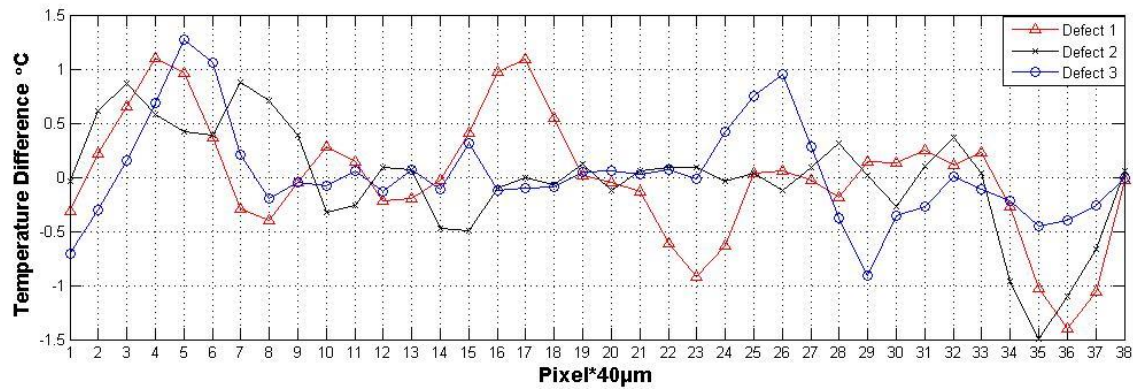


Figure 13. Horizontal temperature difference of defects

Because that the vertical part of the specimen is not all located in the thermal image and in order to better show the boundary of the specimen and the defect, the horizontal temperature difference of the thermal image is used to calculate the defect sizes. Due to the low precision of the infrared thermal imager, it is easy to miss the temperature data of the edge of the specimen and defect. Furthermore, the heating time is slightly larger, and the thermal diffusion is more uniform. Therefore, the selection of pixels should narrow the range. In figure 13, it is clear that the left boundary of the specimen is located on the 2nd pixel and the right is between 34th pixel and 35th pixel. Actually, the size of the specimen width is 5cm. Thus, there are 34 pixels that represent the length of 50 mm. Also, the pixel number of the first defect is 7(16th pixel to 22nd pixel). The second one is about 7(6th pixel to 12th pixel). The third one is about 3-4(25th pixel to 28th pixel). Then, a equation occurs in this section to help for calculating the size of the defect and the relative positions of three defects.

$$S_i = kN$$

$$k = \frac{S_r}{N_0}$$

Where S_r , N_0 , S_i and N_i are the actual size of specimen, number of pixels in the size of the specimen, defect size and number of pixels in the size of the defect, respectively.

It is concluded that the horizontal sizes of the three defects are about 10.09 mm, 10.09 mm and 5.04 mm. And the sizes of relative position of defects are listed in table 3. Based on the result of the error within 5%, it is feasible to use the temperature difference method to quantitatively analyze the relative positions of multiple defects.

Table 3: Relative position of defect

| | Horizontal centre distance | | Vertical centre distance | |
|-----------------------|----------------------------|----------------------|--------------------------|----------------------|
| | (defect 1, defect 2) | (defect 2, defect 3) | (defect 1, defect 2) | (defect 2, defect 3) |
| Actual size/mm | 12 | 22 | 30 | 25 |
| Experimental size/mm | 12.25 | 23.05 | 28.82 | 24.5 |
| Percentage error/100% | 2.1 | 4.8 | 3.93 | 2 |

4. CONCLUSION

A series of experiments were conducted on the carbon steel using the pulsed eddy current thermography and obtained the temperature data of the surface of the sample with multiple defects. According to the signal processing and analysis by the thermal images sequence of valid data, the boundary position and the relative location of the three crack defects can be proposed to detect and quantify defects. The results show that it is feasible to analyze the location of the defects and the relative position of the defects by using the temperature difference method. Furthermore, these will be valuable to improve the defect visibility and widen the pulse eddy thermography detection for metal part with multiple defects.

ACKNOWLEDGMENTS

This research was financially supported by the Tianjin Natural Science Foundation of China (Grant nos. 13JCZDJC34400, 14JCYBJC42100), Hebei Provincial High-level Personnel Funding (Grant nos. E2012100003, E2014100015).

REFERENCES

- [1] An Chunlei, Han Zhennan. Analysis of the influence of pitting and spalling on torsional mesh stiffness of gears[J]. Journal of Vibration, Measurement & Diagnosis. Vol.28 No.4 Dec. 2008.
- [2] Yin A, Cao Bin, Tian Guiyun, et al. Physical interpretation and separation of eddy current pulsed thermography[J]. Journal of Applied Physics, 2013, 113 (6) :064101
- [3] Weekes B, Almond D P, Cawley P, et al. Eddy-current induced thermography-probability of detection study of small fatigue cracks in steel, titanium and nickel-based superalloy [J] . NDT&E International, 2012, 49:47-56.
- [4] Lahiri B B, Bagavathiappan S, Soumya C, et al. Infrared thermography based defect detection in ferromagnetic specimens using low frequency alternating magnetic field induced heating [J]. Infrared Physics & Technology, 2014 , 64:125-133.
- [5] Biju N, Ganesan N, Krishnamurthy C V, et al. Defect sizing simulation studies for the tone-burst eddy current thermography using genetic algorithm based inversion [J]. Journal of Nondestructive Evaluation, 2012, 31 (4) :342-348.
- [6] N. Biju, N. Ganesan, C.V. Krishnamurthy and Krishnan Balasubramaniam, 'Frequency optimization for eddy current thermography', NDT&E International, Vol 42, pp 415-420, February 2009.
- [7] Matthias Noethen, Yi Jia, Norbert Meyendorf. Simulation of the surface crack detection using inductive heated thermography [J]. Nondestructive Testing and Evaluation, 2012, 27(2):139-149.

Cluster-Based Thresholding of Phased Array Ultrasound for Anomaly Detection in Weld Inspection

B. Cassels¹, L.-K. Shark¹, S.J. Mein¹, A. Nixon², R. Turner²

¹ Jost Institute for Tribotechnology, University of Central Lancashire, Preston PR1 2HE, UK

²BAE SYSTEMS – Maritime Submarines

ABSTRACT

In manufacturing and condition monitoring of many engineering structures, the integrity of each weld has to be assessed using a non-destructive test technique. In this context the ultrasonic phased array is a widely accepted method of data acquisition. A linear sequence of sector scans, each created by electronically sweeping the beam of the phased array, provides 3D volumetric data of each weld. Particularly when the engineering structures are physically large this produces a comprehensive but vast amount of data for off-line analysis and has led to the work described in this paper. The aim is to minimise data analysis time through reliable and automatic identification of defective regions. By treating defects as foreground anomalies superimposed on a background base material, the paper presents different cluster-based thresholding techniques that achieve automatic defect detection. In particular, the proposed method first reduces the original 3D data set to 2D using peak values from each beam and image segmentation is then performed to distinguish the background base materials from the foreground defects. The work compares and presents a performance evaluation of the different thresholding methods. In each case robustness is demonstrated and validated based on signal-to-noise ratio using a comprehensive set of ultrasonic phased array data acquired from steel test blocks. These contain different types and sizes of weld flaws at different locations and orientations.

Corresponding author: Bryan Cassels (BCassels@uclan.ac.uk)

1. INTRODUCTION

Ultrasonic echoes from a directed beam of ultrasound provide a means of detecting internal flaws within a weld. The Phased Array (PA), which is both a transmitter and receiver of ultrasound, allows electronic steering of the ultrasonic beam through a range of angles [1]. From a full scan it is possible to create a sector image which represents a cross section through the weld (figure 1.d). For complete data acquisition a number of these sector scans are made at regular intervals along the weld's length with the full data set containing 3D information. At a typical distance between scans of around 2 mm, a full inspection can easily produce many hundreds of sector scans. Each of these must now be visually examined by a qualified inspector. This expensive and error prone stage provides a motivation for automating inspection.

Within the data set interconnected voxels containing unusually high values help to locate anomalous regions. However this is computationally expensive. The proposal here is to reduce this data by taking only the peak values from each beam, thus reducing the 3D volumetric data to a 2D image (figure 2.a). Although some information is lost correlations in the two dimensional data emphasise anomalous regions and the image is more amenable to efficient analysis. Through image segmentation (e.g. figure 2.b) it is possible to locate and, to some extent, size each fault.

Image thresholding is the simplest method of image segmentation. Previous studies [2],[3] have compared thresholding methods for Non Destructive Test (NDT) applications. However these do not specifically relate to ultrasonic PA data from welds. Additionally the ultrasonic data sets are relatively small with fixed Signal-to-Noise Ratio (SNR). In practice a weld may be many metres in length and contain very few defects. Although the test data used here is also limited this issue is addressed by creating virtual test pieces of long welds using ultrasonic data acquired from a small number of representative test blocks. These virtual test pieces can be of any length and can contain any pre-defined set of faults with specified SNR.

This process of creating a virtual test piece also produces a set of fault based test vectors, essentially a list of sectors containing faults, for automatic and statistically based performance analysis.

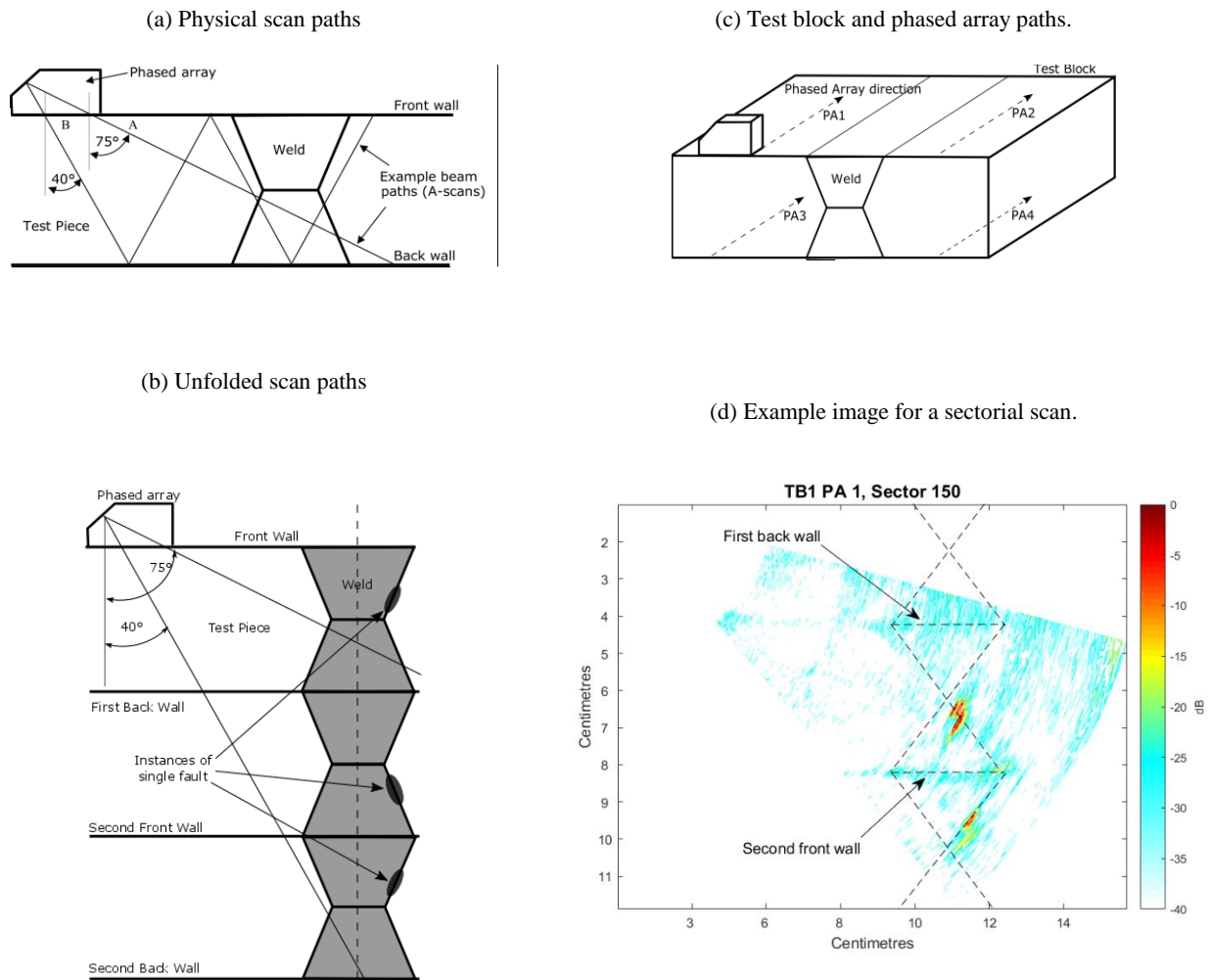


Figure 1. Overview of phased array scanning and the physical set up

Broadly there are three parts to this performance evaluation of anomaly detection: -

- i. The first set of tests evaluates three methods of thresholding and subsequent segmentation of the original test blocks. In each case segmented images are compared, on a pixel-by-pixel basis, to manually generated ground truths.
- ii. The previous threshold values are now used to segment a set of virtual test pieces. The test compares anomalous sectors identified from segmentation with the associated test vectors. As part of this step threshold values are also obtained for the virtual test piece. These are compared to those determined from the original test blocks.
- iii. Finally further evaluation is conducted by means of the Receiver Operating Characteristic (ROC) [4] curve. By scanning through every threshold value (0 to 100%) the ROC curve indicates the effectiveness of thresholding for segmentation; it also allows other classification techniques to be compared. Here the ROC curve provides information on thresholding under different SNRs. It also suggests an optimum threshold value for comparison with those produced earlier.

2. APPROACHES AND METHODS

The test blocks are similar to that illustrated in figure 1(c). Each contains a set of manufactured faults at known locations. The parent material is an austenitic steel and weld lengths are 400 mm. with block thicknesses of 45 mm. The Ultrasonic Phased Array (UPA) allows a beam of ultrasound, an A-scan, to be electronically steered and focussed [1]. As the UPA is manually dragged along the length of the weld at a fixed distance from its centre line, figure 1(c), two sectorial scans (high gain and low gain) are recorded. Each sectorial scan consists of a full set of A-scans, typically through 40° to 75° in 0.5° steps. Figures 1(a) and 1(b) provide a description of a beam's path through the weld. For low angled A-scans (towards 40°) the beam 'skips' through the weld. At higher angles the path is direct.

In figure 1(b) the weld is vertically reflected twice to more clearly illustrate the path of each A-scan. At 40° the A-scan crosses the first back wall then the second front wall before finally crossing the weld's fusion face. The diagram also indicates how a single fault may appear more than once on the corresponding sector scan image, figure 1(d). In some cases three instances of the same fault may be present in the same sector. The use of the test block allows each weld to be scanned four times (PA1, PA2, PA3 and PA4) as indicated in figure 1(c). In practice it may only be possible or desirable to scan the weld once.

In the complete 3D data set anomalies are evident by applying a threshold on voxel intensities. Interconnected voxels above the threshold form regions requiring further investigation. However as a typical A-scan may contain more than 1000 samples a simpler and more efficient procedure is to reduce the data set to 2D by taking only the peak value from each A-scan, figure 2(a). The resulting image is much easier to segment (figure 2(a)).

Establishing an effective threshold value depends on the image content and choice of algorithm. Here Kittler and Illingworth Minimum Error [5] and Maximum Entropy [6] are chosen following the evaluation of Sezgin and Sankur [3]. The Otsu [7] method is added due to its popularity and availability in packages such as Matlab.

In practice the threshold may be determined prior to test and as part of the set-up and calibration procedure. For example using a set of geometrical reflectors (such as side drilled and flat bottomed holes) at different depths and orientations in material similar to that of the weld. This provides precise ground truth information but the faults are artificial.

An alternative, as in this case, is to use a test block with a set of representative manufacture faults. The faults are now more realistic but ground truth information is less precise. Using this method threshold values in directions PA1 to PA4 are obtained for each test block. That is a total of 8 threshold values for each method, table 1.

A severe limitation of the available data is that it is from only two test blocks, additionally the faults tend to have a high signal to noise ratio and are densely packed. This contrasts with the practical situation where welds may be a few metres in length and contain few faults. A more comprehensive investigation would require data from many production welds with a full list of faults, dimensions, locations and orientations prepared by an experienced inspector. This is both costly and not practical. To overcome these limitations the test block data is used to create virtual test pieces. These can be of any length and contain any pre-determined pattern of faults.

Using data from four scan directions (PA1 to PA4) of the two test blocks over 40 fault instances were extracted from the low gain data. The process is semi-automatic. First the fault is manually selected. Ground truth data is taken as the number of sectors the fault occupies. This entails some subjectivity, after some experimenting the sectors chosen were those containing A-scan values greater than -12 dB from the fault's peak plus one sector either side. The outcome of the exercise is the creation of a library containing 41 fault instances. This information provides the ground truth, in terms of a sector count, for each instance.

During a calibration procedure the gain of the low gain channel is set to give a fixed response from a known reflector; the high gain setting is obtained by adding 20 db. While high gain data provides a pronounced response from defects it is often saturated producing an unsatisfactory threshold value that gives many false positives. However, as there is a linear relationship between high gain and low gain background reflections, a fraction of the high gain data can be added to the corresponding low gain background data to simulate an increase in structural noise. Now a virtual test piece of any length and combination of faults can be created by interleaving low gain background sectors (with added noise) and fault sectors from the library.

As details of the virtual test piece are known it is also possible to create a fault based test vector (of faulty sector numbers) to automate the analysis. Table 2 lists 5 test pieces used in the second part of this evaluation. In this case the quoted SNRs are for no added noise. In all cases SNRs are determined using equation 1 where $PEAK_F$ corresponds to the peak value of the fault and RMS_B is the RMS value of the background:-

$$SNR_{fault} = 20 \cdot \log_{10} \left(\frac{PEAK_{fault}}{RMS_B} \right) \quad (1)$$

3. PERFORMANCE EVALUATION BASED ON TEST BLOCKS

Here a threshold value for each direction (PA1 to PA4) of the two test blocks is determined using the three algorithms. These values are then used to segment the corresponding image. Evaluation of algorithms requires a ground truth to which each result may be compared. Test block data sheets are found to give only an approximate location, size and orientation of each implanted fault. Another uncertainty is the precise location of the test probe at the start of the inspection. Establishing a ground truth is therefore a subjective process. For this exercise they are established by segmenting each gray scale image using an increasing threshold level; at some point the features most closely resemble those expected from the data sheet. The example, figure 2(b), gives one illustration. Here the threshold is set so that the majority of foreground (white) areas match those expected from the data sheet.

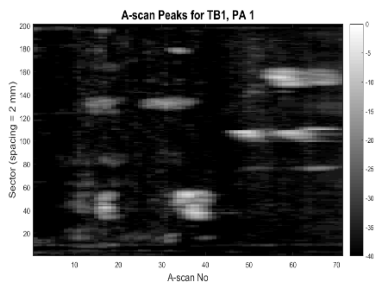


Figure 2(a) Example 2D image

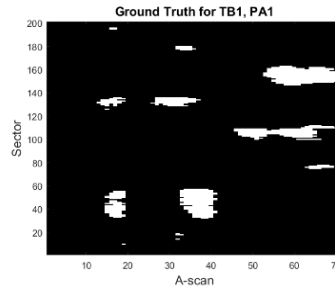


Fig. 2(b) Equivalent ground truth

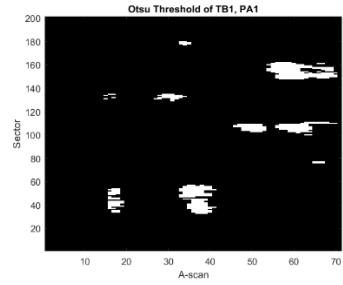


Fig 2© Otsu threshold

Figure 2(c) illustrates the result of applying the Otsu threshold to the original 2D image. In comparison to the manually generated ground truth there is some constriction of the foreground areas indicating that the Otsu threshold value is possibly high. One measure of the difference is the miss classification error [1] which counts the number of foreground pixels erroneously assigned as background (or background erroneously assigned to foreground) with respect to the ground truth image. It is calculated from

$$MCE = 1 - \frac{|B_G \cap B_T| + |F_G \cap F_T|}{|B_G| + |F_G|} \quad (2)$$

Where FG, BG are the number of Foreground and Background pixels of the ground truth and BT and FT represent the number of Background and Foreground pixels of the threshold images.

Figure 3 presents an overview of the performance of each thresholding method in terms of its ability to segment test block data. Results are for PA1 to PA4 of the two test blocks. In all instances the Kittler and Illingworth (K&I) threshold value gives a segmentation most closely matching that of the ground truth.

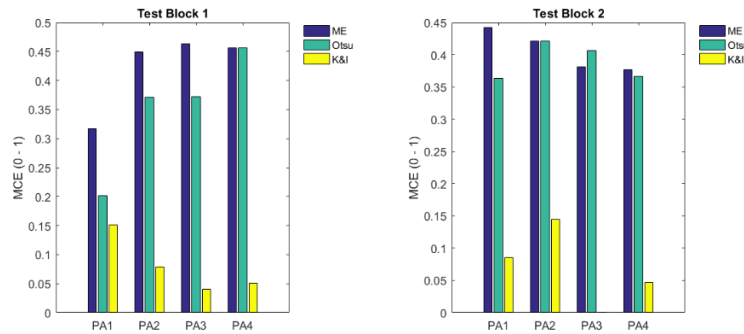


Figure 3. Ground Truth comparison

The threshold values are tabulated in table 1 as a percentage of the total dynamic range. It was expected that, for a given test block, the four threshold values (PA1 to PA4) would be similar. Of the three the K&I approach tends to produce the most consistent results for the same test block; these also tend to be the lowest values.

Table 1: Test Block Threshold Values (%)

| Method | TB1 | | | | TB2 | | | |
|--------|------|------|------|------|------|------|------|------|
| | PA1 | PA2 | PA3 | PA4 | PA1 | PA2 | PA3 | PA4 |
| ME | 12.5 | 15.6 | 18.0 | 18.8 | 14.9 | 18.8 | 12.5 | 11.3 |
| Otsu | 9.4 | 10.9 | 11.7 | 18.8 | 10.1 | 18.8 | 14.9 | 10.9 |
| K&I | 4.1 | 5.2 | 4.1 | 4.1 | 4.5 | 4.9 | 3.7 | 4.1 |

4. PERFORMANCE EVALUATION BASED ON VIRTUAL TEST PIECES

This part of the evaluation uses a set of virtual test pieces (TP1 to TP5), these range in length from 107 to over 2400 sectors representing a weld over 4.8 metres in length. Three faults were selected based on size (number of sectors) and original SNR. When noise is added the resulting SNR is with respect to the smallest fault; in general, the fault covering the smallest number of sectors tends to have the smallest SNR. After segmentation blob detection is used to determine the area (number of pixels) of each feature along with the sectors each occupies. The sectors are compared with those listed in the fault based test vector. Ideally segmentation will identify all sectors containing a fault while omitting all sectors that are fault free. In practice this is unlikely and some measure of performance is needed. These measures are usually in terms of: -

- TP - the number of true positives (sectors correctly identified as containing a fault).
- TN - the number of true negatives (sectors correctly identified as fault free).
- FP - the number of false positives (sectors incorrectly identified as containing a fault).
- FN - the number of false negatives (sectors incorrectly identified as fault free).

One measure is the correct decision rate (alternatively known as the accuracy). This is the ratio of the total number of correct classifications to the total number of decisions ($(TP + TN)/(TP + TN + FP + FN)$).

Using the five virtual test pieces (TP1 to TP5), listed in table 2, the intention is now to evaluate segmentation by applying the previously determined thresholds for TB1 and TB2 (table 1). However before doing this threshold were, for comparison, obtained for each virtual test piece (table 2). Only the K&I thresholds have similarity, indicating that it will segment the smallest fault. There is an increase in the ME

threshold as larger faults are added, suggesting an increase in false negatives. The Otsu method appears to fail completely except where there is a relatively high fault density such as in TP4. The accuracy results, table 2, are for the new thresholds.

Table 2: Accuracy for 5 virtual test pieces with new thresholds.

| Test Piece | Weld Length (sectors.) | Thresholds Virtual Test Piece | | | Fault 1 (small) | | Fault 2 (middle) | | Fault 3 (large) | | Accuracy (%) | | |
|------------|------------------------|-------------------------------|-------|------|-----------------|---------|------------------|---------|-----------------|---------|--------------|----|-----|
| | | KI | ME | O | SNR | sectors | SNR | sectors | SNR | sectors | KI | ME | O |
| TP1 | 107 | 4.5% | 4% | 1.1% | 12.5 | 5 | | | | | 97 | 98 | 5 |
| TP2 | 1407 | 4.5% | 4% | 0.7% | 12.5 | 5 | | | | | 99 | 99 | 0.5 |
| TP3 | 2120 | 4.5% | 4% | 0.7% | 12.5 | 5 | 20 | 12 | | | 96 | 85 | 1 |
| TP4 | 189 | 4.0% | 11.3% | 8% | 12.5 | 5 | 20 | 12 | 29 | 10 | 89 | 93 | 93 |
| TP5 | 2429 | 4.5% | 11.3% | 1% | 12.5 | 5 | 20 | 12 | 29 | 10 | 96 | 95 | 3 |

In the above table TP4 indicates a limitation of accuracy [8] as a means of comparison. For example the values of 93% (ME and Otsu) suggest these thresholds are better than the 85% achieved using K&I. For this work it is essential that a fault is not missed, in fact the K&I threshold of 4.0% segments all faults while the threshold levels of 11.3% and 8% actually exclude the first fault.

Returning to the original thresholds in table 1. The K&I values are in the range 3.7% to 5.2%, those for the Otsu and ME ranging from 15.6% to 18.8%. A selection of these were used to threshold TP5 (the largest virtual test piece). Results, without added noise, are presented in table 3.

Table 3: Accuracy for segmenting TP5 with original Test Block thresholds.

| | Threshold | TP | TN | FP | FN | Accuracy | Comments |
|-------------------|-----------|----|------|-----|----|----------|----------------------------|
| K&I range | 3.7 % | 27 | 1873 | 527 | 2 | 78% | Many FPs |
| | 5.2 % | 27 | 2397 | 3 | 2 | 99.8% | Best accuracy |
| Otsu and ME range | 9.4 % | 17 | 2399 | 1 | 12 | 99.5% | Fault 1 missed |
| | 15.6 % | 13 | 2399 | 1 | 16 | 99.3% | Over 50% of sectors missed |
| | 18.8 % | 10 | 2399 | 1 | 19 | 99.1% | Fault 1 and 2 missed |

The indications are that the K&I thresholds are, for this application the more appropriate. In summary values 3.7% to 5.2% allow identification of all faults; at a threshold of 9.4% fault 1 is no longer detected; at threshold 15.6% over half of the true positive sectors are missed (fault 2 being captured only by 3 sectors). At 18.8% only the largest fault is detected.

5. PERFORMANCE EVALUATION BASED ON ROC CURVES

Further analysis and estimation of performance with respect to added noise can be made if the accuracy measure is replaced by the Receiver Operating Characteristic (ROC) curve [4]. This is a two-dimensional chart in which the true positive rate, or sensitivity ($TP/(TP+FN)$) is plotted against the false positive rate ($1 - TN/(TN+FP)$). The term ($TN/(TN + FP)$) is also known as the specificity (the ability to identify a fault free condition).

An example ROC curve is presented in figure 4. A value for the Area Under Curve (AUC) approaching 1 indicates a good separation between feature and background. If the area approaches 0.5 (represented by a diagonal line from the origin to co-ordinate 1,1) the ability to separate the two becomes useless. The usual purpose of a ROC curve is to compare the performance of two or more classification techniques. In this case it is used to help evaluate thresholding under different noise conditions. In the example threshold values in the range 0 to 100% are applied to a test piece with a fixed SNR. For each threshold value a corresponding *sensitivity* and *1-specificity* point is plotted. This is then repeated after adding noise to the same test piece. With an SNR of 12 dB the AUC is 0.92 while with an SNR of 9 dB it reduces to 0.72.

Table 4 lists the AUCs for each virtual test piece for different SNRs. As expected increasing the noise level compromises segmentation.

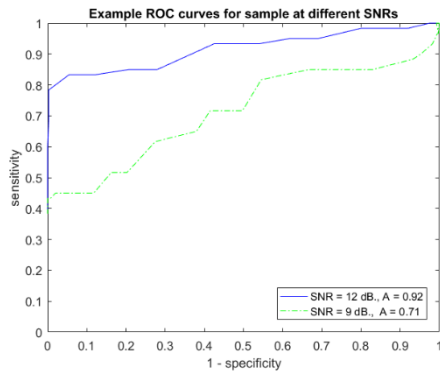


Figure 4. Example ROC curves

Table 4: ROC Areas

| Peak SNR (dB) | 10.5 | 11 | 11.5 | 12 | 12.5 |
|---------------|------|------|------|------|------|
| TP1 (AUC) | 0.76 | 0.78 | 0.8 | 0.85 | 0.87 |
| TP2 (AUC) | 0.78 | 0.82 | 0.83 | 0.86 | 0.88 |
| TP3 (AUC) | 0.91 | 0.93 | 0.94 | 0.95 | 0.96 |
| TP4 (AUC) | 0.91 | 0.93 | 0.93 | 0.94 | 0.95 |
| TP5 (AUC) | 0.92 | 0.94 | 0.95 | 0.96 | 0.96 |

The ROC curve may provide further guidance with threshold evaluation. The suggestion is that, in these examples, the optimal threshold value will correspond to the point on the curve that is closest to the top left corner ($sensitivity = 1$, $1 - specificity = 0$) or: -

$$th_{opt} = \min \sqrt{(1 - sensitivity_{th})^2 + (1 - specificity_{th})^2} \quad (3)$$

The resulting optimal thresholds for the original five test pieces are found to be in the tight range of 3.9% to 4.7%. As the SNR is reduced then at around 11 dB they converge towards 4.9% for each test piece. These results are consistent with the K&I threshold values obtained in table 1 and table 2.

The main requirement of image segmentation is to identify all anomalies. Some false positives can be tolerated but if the number is large the technique becomes useless. The ROC analysis suggests that the threshold values produced by the K&I method provides a good compromise.

Table 5 gives results of thresholding a test piece containing 6 faults with threshold values of 3.9%, 4.7% (corresponding approximately to the range of K&I values for TB1 and TB2 in table 1) and 9.4% (corresponding to the lowest Otsu value). Noise has been added to the test piece to reduce the SNR of the smallest fault to 11.5 dB. The 'Fault:Sectors' column lists the actual number of sectors occupied by each fault. For each threshold level (Th) the corresponding 'Sectors' column gives the number of (true positive) sectors identified by the test bench. The 'Area' column indicates the size, in pixels, of each foreground region in the segmented image. Where the 'Area' column contains more than one number this indicates either more than one instance of the fault in the sector or the fault has segmented into two due to a high threshold value.

Table 5: Test Bench Results (SNR = 11.5 dB)

| Fault | | Th = 3.9% | | Th = 4.7% | | Th = 9.4% | |
|-----------------|---------|-----------|---------|-----------|---------|--------------|---------|
| No | Sectors | Area | Sectors | area | Sectors | area | Sectors |
| 1 | 5 | 9,7 | 3 | 5,4 | 3 | Not detected | |
| 2 | 9 | 33,25,24 | 6 | 29,24,18 | 5 | 6,4 | 2 |
| 3 | 14 | 286,61 | 14 | 266, 55 | 14 | 118,43 | 14 |
| 4 | 7 | 48,33,20 | 6 | 33, 32,16 | 5 | 12,2 | 3 |
| 5 | 12 | 76,74,68 | 11 | 72,69,60 | 10 | 11 | 5 |
| 6 | 23 | 186,61,9 | 23 | 173,53,5 | 23 | 97,3 | 20 |
| False Positives | | 203 | | 57 | | 4 | |

For the threshold values of 3.9% and 4.7% all faults are identified. At a threshold of 9.4% fault 1 is no longer detected. To be of practical use the approach must minimise the number of false positives. As expected these reduce as the threshold is increased. However the size and number of true positive sectors reduce, making the detection more marginal. For a threshold of 4.7%, 57 false positives are listed, however

many of these have a size of 3 pixels or less; in all they account for 44 of the false positives. It is unlikely that these indicate faults giving some scope for further post-processing. This might, for example, include locating the centre of the potential anomaly within the original 3D volume. From the intensity values of the interconnected voxels a 'mass' value for the anomaly could be used for further discrimination.

6. CONCLUSIONS AND FUTURE WORK

This work has investigated the use of image segmentation to identify anomalies in a sequence of sector scans taken from welded austenitic steel blocks. In practice such inspections produce a large amount of 3D data that is time consuming to inspect even by computer. By extracting peak values from each A-scan the resulting 2D image is far easier to process. Thresholding is now the simplest way to detect anomalies and the reliability of this approach has been investigated. Determining a threshold value is a critical. This can be done using a reference block at the discretion of an inspector or automatically using a thresholding technique. In the latter case the K&I method of thresholding has proved to be more reliable than the ME or Otsu methods.

A major obstacle is that a reference block will not always be available. Under this circumstance it becomes desirable to automatically detect regions that are background without any prior knowledge (other than the constraint that the weld geometry is constant). In the case of background A-scans at the same angle and across consecutive sectors take the same path through the weld and have a similar pattern. Any variations are statistically small and occur at around specific samples points (for example those corresponding to the fusion face or a weld cap). Providing faults are infrequent it is possible to 'learn' the background pattern and detect any anomaly.

This type of problem is not unique to NDT. For anomaly detection more generally the use of Robust Principal Component Analysis (RPCA) [9] is receiving attention. This is an extension to conventional PCA which itself shares some of the limitations described here. However RPCA has the potential to extract normal behaviour (background) statistics even when the training data itself contains outliers (anomalies). A condition of this being that, during training, outliers are sparse. The expectation is that most welds are in this category (i.e. they are mainly fault free). The application of RPCA is the subject of further work in this area.

REFERENCES

- [1] Von Ramm O., Smith S, "Beam Steering with Linear Arrays", IEEE Transactions on Biomedical Engineering, vol. BME-30, Issue: 8, 1983.
- [2] R. Saranya, J. Daniel, N. Chermakani, "Comparison of segmentation techniques for detection of defects in non-destructive testing images", 2014 International Conference on Electronics and Communication Systems (ICECS), Coimbatore, 2014, pp. 1-6.
- [3] M. Sezgin, B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation", Journal of Electronic Imaging, Vol. 13, No1 pp. 146-156, 2004.
- [4] T. Fawcett, "An introduction to ROC analysis", Pattern Recognition Letters, vol. 27, no 8, pp 861-874, June 2006.
- [5] J. Kittler, J. Illingworth, "Minimum Error Thresholding", Pattern Recognition, Vol. 29, pp 273-285, 1986.
- [6] J. N. Kapur, et al, "A new Method for Grey-level Picture Thresholding Using the Entropy of the Histogram", Computer Vision Graphics and Image Processing, vol. 29, pp 273-285, 1985.
- [7] N. Otsu. "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber, vol. 9, no 1, pp 62-66, 1979.
- [8] F. Provost, T. Fawcett, R. Kohavi, "The Case Against Accuracy Estimation for Comparing Induction Algorithms", Proceedings of the Fifteenth International Conference on Machine Learning, pp 445-453, 1997.
- [9] E., J. Candes, Xiao Dong Li, Yi Ma, J. Wright, "Robust Principal Component Analysis?", Journal of the ACM, vol. 58, pp 11:1-11:37, June 2011.

A Novel Two-Stage Approach to On-Site Condition Monitoring

V. Leavers

V4L Particles Ltd. 59-60 Thames Street, Windsor SL4 1TX UK

ABSTRACT

This paper sets out an innovative, two-stage approach to wear debris particle analysis for on-site equipment health monitoring. The approach takes account of the fact that conventional laboratory oil testing has recently become less useful for timely condition monitoring information due to the advent of improved lubricant formulations and more efficient filter designs. The two-stage approach uses the latest advances in on-site imaging hardware combined with software that contains an extensively researched knowledge base, which has captured the diagnostic skills and experience of a number of expert wear debris analysts, each with decades of hands-on experience. First stage testing uses the latest computer vision technology to visualise fine debris. This enables on-site maintenance professionals not only to reliably size and count particles but also to analyse fine wear debris as small as 5 microns, offering timely equipment health information that few specialist laboratories can match. The second stage consists of on-site, in-service filter analysis triggered by the appearance of abnormalities in the fine debris particles during first stage analyses. Innovative, ASTM D7684-11 compliant, diagnostic wear debris particle analysis software then uses the five level severity rating advocated by the standard guide such that timely, graded alerts do not allow wear to escalate to a critical level. The paper includes the results of a forensic case study illustrating the way in which a catastrophic bearing failure, costing millions of euros in critical equipment down-time could easily have been avoided had the two stage condition monitoring methodology been applied.

Keywords: Tribology, Condition Monitoring, Wear Debris Analysis.

Corresponding author: V. Leavers v.leavers@V4L-group.co.uk

1. INTRODUCTION

This paper sets out an innovative, two-stage knowledge-based approach to wear debris analysis for on-site equipment health monitoring. The need for a two stage approach arises because conventional laboratory oil testing has become less useful for machine health monitoring. This is due to the advent of improved lubricant formulations and more efficient filter designs, which either remove or significantly delay the appearance of standard, data-driven condition monitoring alert signals from in-service lubricant samples. First stage routine monitoring is of microscopic debris and uses the latest computer vision technology[1] to analyse fine debris in order to extract previously inaccessible, important visual cues from images of particles as small as 5 microns. Abnormalities in the fine debris then trigger early warning alerts to activate the second stage analysis.

Second stage diagnostic analysis is of macroscopic debris extracted from in-service filters or magnetic plugs. This uses interactive software[2] with a knowledge-base for wear debris identification and the associated five level severity rating advocated by the ASTM D7684-11 standard guide[3] to inform timely and appropriate maintenance action, which can help to avoid costly false alarms or critical unscheduled down-time.

2. THE NEED FOR TWO-STAGE ON-SITE CONDITION MONITORING

Technological advances in lubricant formulations and more efficient filter designs should logically have made equipment health monitoring easier. However, this is not the case if monitoring is restricted to conventional oil and debris analysis methods. This is because:

1. Innovative high-performance additives and improved synthetic base oils are more chemically stable and robust. Provided they are used properly and not allowed to become contaminated in storage or use, they should give reliable service. This means that, for most machines, standard laboratory oil tests are only useful to confirm that: the equipment was filled with the correct grade of oil; the correct grade of oil was used for top-ups; and the oil left in the machinery has not become contaminated and/or exhausted.
2. The end result is that modern in-service lubricants are now less likely to reveal machine wear problems when tested by standard laboratory methods such as viscosity and TAN/TBN measurements etc.. This is because the new lubricants are chemically more stable and are less affected by overheating or abnormal wear, than previous lubricant formulations.
3. Despite better lubricants, wear still occurs and wear debris particle monitoring remains the best way to detect faults in lubricated machinery. However, improved filter designs now trap large wear and contaminant particles more efficiently, leaving fewer in the lubricant to analyse by methods such as particle sizing and counting and ferrography.

3. TWO-STAGE WEAR DEBRIS ANALYSIS

To-date it has been the accepted wisdom that it is not possible to extract condition monitoring information from images of particles smaller than 20 microns other than to size and count them. However, dedicated research has produced advances in computer vision technology[1] that now allow important visual information to be retrieved from images of microscopic particles as small as 5 microns.

3.1. First stage testing

First stage testing consists of analysing microscopic debris using state-of-the-art computer vision imaging technology [1]. In addition, automatic particle sizing and counting software has been developed, which is uniquely 'plug-and-play' and does not require the user to input subjective, image-processing thresholds in order to distinguish particles from the background image. This makes it ideal for on-site use where the end user may not have the skill or training necessary to set such thresholds. The new particle sizing and counting hardware and software technology is compliant with the ISO 4406 and 4407, NAS 1638 standards and also the SAE ARP598 standard.

The new technology allows previously inaccessible, important visual information to be extracted from images of particles as small as 5 microns. The software does this by detecting and amplifying important visual cues such as colour and shape. Detectable abnormalities might include: a significant increase in the number of particles <6 microns coupled with the appearance of corrosion particles; evidence of contamination by mineral particles coupled with fine cutting wear or the appearance of fine temper coloured particles.

3.2. Second stage testing

Once the second stage analysis is triggered by abnormalities in the fine, microscopic wear debris the in-service filter or magnetic plug is removed and the larger, macroscopic wear debris particles are extracted for further analysis using software based on an innovative new concept in wear debris particle analysis developed specifically to meet the needs of on-site technicians[2]. The software is compliant with and uses the particle classifications and nomenclature given in the ASTM D7684-11 Standard Guide for the microscopic characterisation of particles from in-service lubricants [3].

The software provides the on-site maintenance professional with access to an expert-knowledge-base of the fundamentals of wear debris analysis in order to assist in the identification of transitions between benign, active and critical wear patterns. It is ideally suited for on-site situations where the level of training and skill of the attendant technician may require substantial support. The user is taken from the particle image to the relevant information without the need to first interpret complex parametric data relating to particle morphology. This ASTM D7684-11 compliant, diagnostic wear debris particle analysis software also uses

the five level severity rating advocated by the standard guide such that timely, graded alerts do not allow wear to escalate to a critical level.

3. FORENSIC ANALYSIS IN THE CASE OF A CATASTROPHIC BEARING FAILURE

A forensic check was carried out in the case of a catastrophic bearing failure due to misalignment and serves to illustrate that the two-stage approach would have detected the fault at a stage when it could easily have been remedied.

Misalignment in rolling element bearings causes uneven cyclic loading which can exceed the design parameters of the bearing components and materials, eventually leading to failure. Initially the cyclic loads caused by misalignment may be insufficient to break the oil film. They are transmitted hydraulically to the metal surfaces, causing fatigue micro-cracking and spalling of the inner and/or outer rings and the balls/rollers. The particles produced by the fatigue wear vary in size and number according to the size and speed of the bearing; the load; the metallurgy of the materials; and the type and extent of the misalignment. However, even at this early stage microscopic particles (5-10 microns) showing evidence of tempering will be generated and these were seen in the archived filter patches routinely created in order to size and count particles in the sampled lubricant. While the sizing and counting results did not exceed the alarm limits with respect to the specified ISO 4406 cleanliness code, microscopic tempered particles had already begun to be generated. Analysis of a time series of archived filter patches showed that the percentage of tempered particles in the samples increased exponentially while the size and count results and the standard TAN and viscosity tests did not exceed their alert levels.

Had the two stage-approach been used the continued increase in microscopic tempered particles would have signalled a need to remove and check the filter. That is, once tempered particles are detected the lubricant should be sampled more frequently and a note kept of the percentage of tempered particles in the sample. If the fault is temporary the percentage of tempered particles may reduce or simply not increase over time. However, if the fault is on-going the percentage of tempered particles will cascade and increase exponentially. This is the signal to remove the filter and begin the second-stage analysis of wear debris particles.

In this case, examination of the filter revealed that some of the macroscopic particles showed direct evidence of fatigue micro-cracking, while others were reworked and flattened by entrainment in the rolling contact of the bearing. Reworking causes local overloading and further fatigue damage to the bearing surfaces. In this case the damage had been left unchecked and had cascaded. Surface scarring of the rings and balls/rollers had disrupted the oil flow within the bearing leading to loss of lubrication, metal-to-metal contact and adhesive wear. Damage to the cage had also occurred. The bearing had overheated and locked up. The two-stage approach would have produced a maintenance action alert that could have halted the damage caused by misalignment at an early stage and before the fault had caused a catastrophic failure.

4. CONCLUSIONS

Because the two-stage approach takes account of the step change in lubricant formulations and filter design it meets the various challenges posed by on-site condition monitoring as detailed below

Advanced wear debris analysis using state-of-the-art digital imaging hardware and software can distinguish temporary and on-going wear situations in a way that conventional oil monitoring tests such as particle sizing and counting, viscosity and TAN measurements cannot. For example tempered steel particles are symptomatic of lubrication starvation, misalignment or heavy loading. The onset of these conditions can be seen in the fine debris particles before it is evident in the standard TAN or viscosity tests. This is because local wear conditions do not usually generate sufficient heat to oxidise the bulk of the lubricant and will therefore not be detectable by standard laboratory tests until the condition is well advanced. Trending of the percentage of fine tempered particles can help to distinguish between transient and on-going wear conditions.

Standard laboratory oil testing produces sets of numbers that may be difficult to interpret in relation to maintenance action. Moreover, changes in the various oil test results may lag behind the onset of the wear condition. This can lead to a simple stop/go maintenance strategy in which costly false alarms are unavoidable. The two stage approach uses visual cues associated with both the microscopic and macroscopic wear particle debris to determine the root cause of the problem and this informs timely and appropriate maintenance action. This can help to avoid costly false alarms when equipment is taken out of service before it is necessary or perhaps more importantly stop equipment from being left in-service when it is critical, which in some industries can lead to fatalities.

When large, critical equipment situated in remote, harsh environments fails unexpectedly it can take many weeks to arrange to repair or replace it. For this reason advance warning of necessary maintenance or replacement can offer significant savings and avoid the losses incurred by avoidable, lengthy deferred revenue situations due to equipment unexpectedly needing to be taken out of service. The two-stage approach can help to determine not only the root cause of an equipment problem but also the appropriate maintenance action according to the five level severity rating advocated by the ASTM D7684-11 standard guide. Thus, avoiding the heavy losses associated with unscheduled critical equipment down time.

Maintenance professionals are able to carry out the checks needed in the two-stage approach on-site. This is the ideal place for equipment health monitoring as engineers can use their extensive knowledge of the equipment to put condition monitoring data and information into the context of maintenance schedules and real-time data: speeds, loads, temperature and oil pressure or unusual operating conditions. Use of the two-stage approach and state-of-the-art digital imaging equipment combined with interactive software support offers a cost effective on-site diagnostic capability to rival that of most specialist labs.

REFERENCES

- [1] Leavers V. Advances in Wear Debris Analysis with Relevance to the Needs of On-Site Condition Monitoring, (2013) Proceedings OilDoc Conference, Rosenheim Germany 2013.
- [2] Leavers V., (2014) Collection, Representation and Retrieval of Expert Knowledge Related to Wear Debris Mode Classification, Proceedings of LubMat 2014
- [3] ASTM D7684-11, Standard Guide for Microscopic Characterisation of Particles from In-Service Lubricants

Session 3

Monitoring, Diagnosis, Prognosis and Health Management 1

Session Chair Prof. Mike Barnes

Wind Turbine Fault Detection by Monitoring its Performance Using High-Resolution SCADA Data

(E. Gonzalez, J.Melero)

Estimation Method of Static Contact Wire Height by Using Contact Force on Commercial Line

(T. Usuda and M. Ikeda)

Feature Selection for Remaining Useful Life Prediction of Spur-Bearings

(K. de Calle, S. Ferreiro, E. Konde, I. Bravo, A. Arnaiz)

Active Lubricant Condition Monitoring *(J. George, A. Onsy)*

Value Based Management in Centrifugal Compressors Maintenance

(X. Liang, F Duan, D. Mba, I Bennett)

Wind Turbine Fault Detection by Monitoring its Performance Using High-Resolution SCADA Data

E. González¹, J. J. Melero¹

¹ CIRCE-Universidad de Zaragoza, C/Mariano Esquillor Gomez, 50018 Zaragoza, Spain

ABSTRACT

Much effort is recently dedicated to develop condition monitoring system (CMS) for wind turbines (WT) utilising the data of the Supervisory Control and Data Acquisition (SCADA) system. Performance monitoring is a fundamental task in the development of CMS for WTs. Component malfunction can influence the performance, resulting in significant deviations. The most widely approach is monitoring WT performance through its power curve. WT power curve can be modelled either with parametric and non-parametric approaches; control charts based on residual monitoring are a useful tool to detect performance deviations. While many contributions have been made in WT power curve modelling, no correlation between performance deviations with real component failure data has been studied. This paper aims to address WT fault detection capabilities of performance monitoring techniques. First, state-of-the-art machine learning techniques are used to model WT power curve based on high-frequency SCADA data. Indeed, most of the power curve modelling techniques rely on the use of 10-minute data, that may smooth dynamic effects. Then, performance deviations are categorised by faulty WT component by using the SCADA alarm system. Finally, fault detection capabilities are proved by comparing detected deviations in healthy and faulty wind turbines as gathered in suitably documented failure information.

Keywords: Wind turbine, Operation and Maintenance, Performance Monitoring, SCADA, Fault detection

Corresponding author: E. Gonzalez (email: egonzalez@fcirce.es)

1. INTRODUCTION

Wind power installed capacity accounted for 51% of total installations in 2016, being the main source of installed power among generation facilities in Europe [1]. Although some countries have seen a notable slowdown, the sector is still growing with new onshore installations and wind turbines (WT) of increasing size; furthermore, offshore wind has experienced remarkable growth rates over the last years. This large scale-deployment has brought challenges to the operation and maintenance (O&M) of wind farms. WT failures together with poor accessibility contribute to high O&M costs and have a strong impact on downtime and thus on annual energy production (AEP). This is especially critical for offshore wind farms, as shown in [2]. Predictive maintenance is a promising approach for reducing the O&M costs; early detection of failures not only contributes to better maintenance planning but also to increased turbine availability and AEP. Condition monitoring is a tool commonly employed for the early detection of WT failures. Many condition monitoring systems (CMS) developed for WTs are component-specific and usually require the installation of costly additional equipment [3]. Alternatives solely based on the use of data from the supervisory control and data acquisition (SCADA) system are very promising as these data are readily available on modern WTs. Since failures of critical components may negatively affect WT performance, monitoring it is an important step in SCADA-based condition monitoring. Deviations from faultless performance might be used as indicators of upcoming failures.

Performance monitoring is generally synonymous with monitoring WT power curve. A reference power curve should capture WT performance under normal operation and it can thus be used to detect deviations. However, building this reference is not straightforward. Using the ideal power curve provided by a wind

turbine manufacturer is not possible, since it is only warranted under specific conditions. In reality, on-site conditions may be quite different from those stipulated in the turbine supplier agreement (TSA), and power curves are greatly influenced by several phenomena, e.g. atmospheric stability [4], wind shear and turbulence intensity [5], but also by control strategies and component conditions. Data-driven methodology is a widely used approach to model WT power curve since it allows to take into account any feature inherent in the data.

Many approaches for WT power curve modelling can be found in the literature, especially data-driven, based on data mining and machine learning algorithms. These can be classified as either parametric or non-parametric algorithms. Some parametric examples are polynomial regressions [6], [7], logistic functions [8], [9], linearized segmented model [9], the linear Hinge model [10] and the modified hyperbolic tangent [11]. While parametric models can be summarised as a function that maps input variables to output variables, non-parametric algorithms do not make prior assumptions about this mapping function. These have been widely applied to power curve modelling. Kusiak et al. were first to apply k-nearest neighbour (k-NN) in [8], used subsequently in [12], [13]. Many contributions have been made investigating neural networks (NN). Self-organised map (SOM) and artificial neural networks were used to detect performance degradation [14]. Underperformance was also detected in [15] by the use of auto-associative neural networks (AANN). Other examples can be found in [9], [12], [16]–[18]. Data-mining algorithms were also investigated by Lidia et al in [9]; an interesting comparison between parametric and non-parametric algorithms for power curve modelling was shortly after published by the same authors [19]. Kernel methods were explored in [20] and recently in [21], where Ouyang et al successfully built a power curve by the use of support vector machines. Non-parametric models have also been used to monitor the whole wind farm power [22]–[24]. Gaussian process [17], [25], [26] or copula modelling [27] are interesting approaches to obtain confidence levels for power curve monitoring. Finally, although it is a complex process influenced by several phenomena, few examples can be found involving multivariate power curve modelling [28], [29], suggesting an improvement in modelling accuracy.

In general, non-parametric models are more accurate than parametric. Indeed, a power curve should reflect a chain of complex conversion processes, which moreover varies according to different operating regimes. Nevertheless, while much effort has been devoted to power curve modelling and monitoring, not many contributions discussed the reason for power deviations, particularly related to component status. Degradation was detected prior to a gearbox failure in [6] and deviations were observed during a generator over-temperature fault in [26]. This is undoubtedly due to the lack of failure data - access to datasets is often proprietary. Furthermore, most of the contributions used 10-minute aggregation SCADA data as it is an industry standard. We believe this aggregation may hide rapid events, as stated in [12]; new high-resolution SCADA systems should allow to represent dynamic turbine behaviour with higher fidelity and therefore improve detection capabilities. In this paper, high-resolution SCADA data are used together with SCADA logs and suitably documented failure information to develop a methodology for WT fault detection by monitoring its performance. For this purpose, multi-variate power curve modelling techniques are applied. The performance of a healthy turbine and a turbine affected by a gearbox failure are compared.

The paper is organised as follows. Section 2 describes the whole proposed approach. Section 2.1 covers data description and pre-processing methodology and section 2.2 discusses modelling and uncertainty assessment for monitoring purposes. Section 3 discusses accuracy of the different suggested models. In section 4, a case study is presented, where performance is monitored for two turbines, one faulty one healthy, in a concurrent period before a failure. Finally, conclusions can be found in section 5.

2. METHODOLOGY

The framework developed integrates both operational and alarm log SCADA data to produce a multivariate power curve model together with a related uncertainty, to build a normal performance reference. Actual power observations outside normal performance are detected as events of underperformance and

categorised by component or issue through the use of alarm logs. A general sketch of the methodology is shown in figure 1.

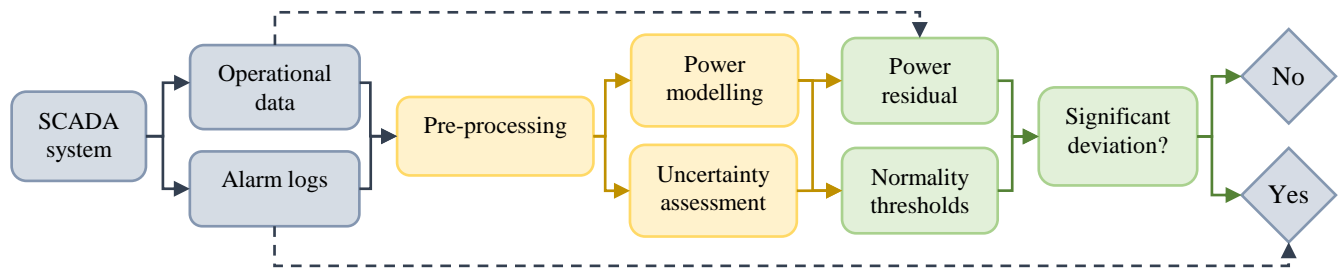


Figure 1. Suggested framework for wind turbine performance monitoring.

2.1. Data description and pre-processing

The SCADA data used in this paper comes from a commercial wind farm; data from turbines H (*healthy*) and F (*faulty*) were used in this work. The two turbines, located next to each other, should be exposed to similar local conditions. Three full months of high-resolution data were collected: the first month was used to train models in both turbines, while performance was monitored during the two other following months. Dates were chosen as WTF exhibits a gearbox failure on 21/03/2015. A description is presented in table 1.

Table 1: Description of the datasets

| Dataset | Duration | Sampling interval | Description |
|---------|-------------------------|-------------------|---------------------------------------|
| WTH - 1 | 01/01/2015 – 01/02/2015 | 4-second | Training data – 146090 observations |
| WTH - 2 | 01/02/2015 – 01/04/2015 | 4-second | Monitoring data – 260440 observations |
| WTF - 3 | 01/01/2015 – 01/02/2015 | 4-second | Training data – 99107 observations |
| WTF - 4 | 01/02/2015 – 01/04/2015 | 4-second | Monitoring data – 256453 observations |

Operational data, i.e. nacelle wind speed, air temperature, pitch angle, rotor speed and active power, were flagged by the integration of the alarm logs, translated into component-related information. The SCADA logs are records of relevant information or issues affecting each wind turbine. An alarm is activated when an incident occurs or to keep records of any operational status change. Based on a modernised wind turbine taxonomy [30] and original technical documentation from wind turbine manufacturers, each unique alarm code was manually associated to a specific *sub-system* and *assembly*. Each operational data record can be therefore related to an affected component or a specific issue. More information can be found in research previously published by the authors [31]. The data registered while wind turbine was experiencing an issue were discarded during the pre-processing phase. While some previous work keep pre-processing to a minimum [29], we believe this step is crucial for power curve modelling accuracy. Data flagged as abnormal can be seen in figure 2. All figures presented here are normalised for confidentiality reasons.

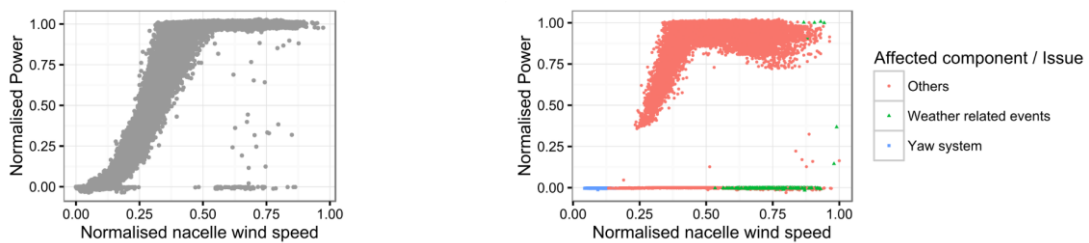


Figure 2. WT normal vs. abnormal performance as flagged by component-related logs.

To ensure faultless representation of WT operation, further abnormal performance filtering was applied based on multivariate curve approach and machine characteristics. The understanding of WT performance is highly improved as one can see in figure 3. Pitch angle and rotor speed allow to clearly identify the different operating regimes.

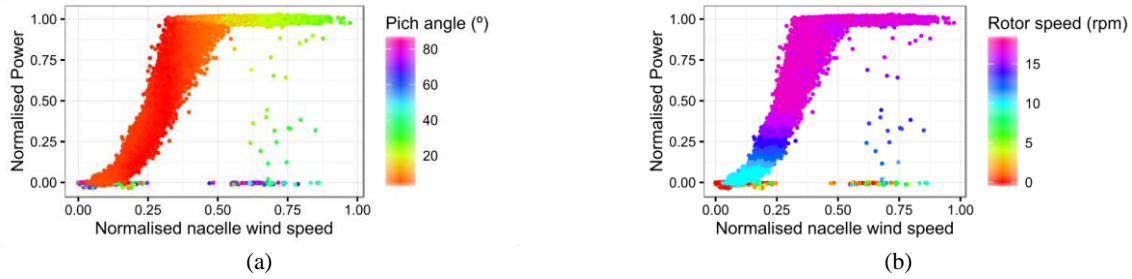


Figure 3. WT power curve varying depending on blade pitch angle (a) and rotor speed (b).

It has to be mentioned that the variables used in this work exhibit an unevenly distribution over time. Data registration often depends on the SCADA system configuration; while wind speed and power are recorded every 4-seconds, air temperature is only updated every 10 minutes. To overcome this heterogeneity, data were resampled by linear interpolation between consecutive observations. This assumption can be made due to the low rate of change of the air temperature.

2.2. Performance modelling and uncertainty assessment

In this section, we characterise WT performance by multivariate power curve modelling. In all, six models were trained and tested by using a 10-fold cross-validation, by randomly dividing the observations into ten-folds of approximately equal sizes. Uncertainty related to predicted power was also assessed for each model. For monitoring purposes, it is important to build a threshold of normal operation that should account for prediction variability. The tested models are described subsequently.

- **M1-Method of bins (MOB)**

The method of bins is a data reduction procedure defined in the standard IEC 61400-12-1 [32] to measure the power curve of a single wind turbine, used as a characteristic of its performance. It relies on the reduction of the dataset into mean values per wind speed intervals, called bins. For each bin, centred on a multiple of 0,5 m/s, the wind speed and power output are reduced to the calculated mean, that is:

$$V_i = \frac{1}{N_i} \sum_{j=1}^{N_i} V_{n,i,j} \text{ and } P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{n,i,j}$$

Each bin is therefore represented by a vector of wind speed and power output. In this paper, the predicted power was obtained by applying the reference power curve to the nacelle wind speed, where the reference power curve was built through a linear interpolation between bins. Normal performance thresholds can be built based on the standard deviation of power in each bin.

- **M2-Generalised Linear Models (GLM)**

GLMs are an extension of linear modelling framework to response variables that are not normally distributed; they can therefore be used to model continuous, ordered and unordered data. Since an underlying linear relation between explanatory variables and response may not be the same for the whole power curve, data were partitioned according to the different operating regimes. Being y the power, x_1 the wind speed, x_2 the pitch angle, x_3 the rotor speed and x_4 the air temperature, five different regions were identified, as one can see in table 2. The definition of the regions was done according to the changing rates of variation of the pitch angle and rotor speed. Concerning uncertainty assessment for normality thresholds, confidence intervals can be built based on the prediction and the standard error of the prediction (standard error of the fitting).

Table 2: Description of the data partitioning for M2 training

| | $\frac{\partial y}{\partial x_1}$ | $\frac{\partial y}{\partial x_2}$ | $\frac{\partial y}{\partial x_3}$ | $\frac{\partial y}{\partial x_4}$ |
|----------|----------------------------------------|-----------------------------------------|-----------------------------------|-----------------------------------|
| Region 1 | ≈ 0 | ≈ 0 | ≈ 0 | — |
| Region 2 | ≈ 0 | < 0 | Fixed rotor | — |
| Region 3 | ≈ 0 | $\approx 0, \leq 0$ | > 0 | — |
| Region 4 | ≈ 0 | $\approx 0, \geq 0$ | Fixed rotor | — |
| Region 5 | ≈ 0 (Increasing wind speed) | ≈ 0 (Increasing pitch angle) | ≈ 0 Fixed rotor | — |

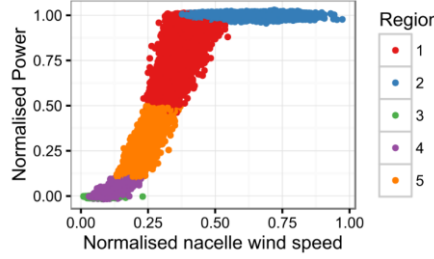


Figure 4. Data partitioning for training of M2.

- **Non-parametric models.**

Four non-parametric models were selected based on the literature (see section 1). These are k-nearest neighbour (M3 - kNN), regression trees (M4 - RT), random forest (M5 - RF) and support vector regression (M6 - SVR). Detailed information about these algorithms can be found in [33].

The kNN [8], [12], [29] approach predicts a new sample using the k closest samples from the training dataset; the selected distance between samples is the Euclidean distance. The predicted response for a new observation is the mean of the k nearest responses in the training dataset. The parameter k ($k = 22$) was selected in order to minimise the testing error. RTs [29], [34], [35] partition the data into smaller groups that are more homogenous with respect to the response. The algorithm determines the predictor to split and its value, the depth or complexity of the tree and the prediction equation in the terminal nodes, that can be simply the average of the training set outcomes in that node for prediction. The methodology applied here is based in the well-known classification and regression tree (CART) [36]. While in standard trees, each node is split using the best split among all variables, in RFs [8], [29] each node is split using the best among a subset of predictors randomly chosen at that node. Detailed information can be found in [37]. Finally, SVR [8], [21], [38] is the unique kernel method applied here. Kernel methods require a specified kernel, i.e., a similarity function over pairs of data points in raw representation. Radial Basis kernel (Gaussian) was selected here as no prior knowledge is supposed about the data.

Uncertainty related to prediction was assessed similarly for each method, based on the residuals from the training dataset, as stated in [8]. Normal performance thresholds were built based on the mean residual (μ_{train}) and the standard deviation (σ_{train}), as shown in the following equations. The parameter η was selected in order that 95% of the training residuals were retained within the normal performance thresholds.

$$\begin{aligned} \text{Upper Normal Performance} &= \mu_{train} + \eta \frac{\sigma_{train}}{\sqrt{N_{train}}} \\ \text{Lower Normal Performance} &= \mu_{train} - \eta \frac{\sigma_{train}}{\sqrt{N_{train}}} \end{aligned}$$

This approach is valid in the case of the homoscedasticity; however, both the power values or the residuals between predictions and actual observations show a clear heteroscedasticity. Therefore, the way normal performance thresholds are calculated is not suitable for our case. To overcome this issue,

data were partitioned again, creating different thresholds for the different operational regimes, as one can see in figure 5.

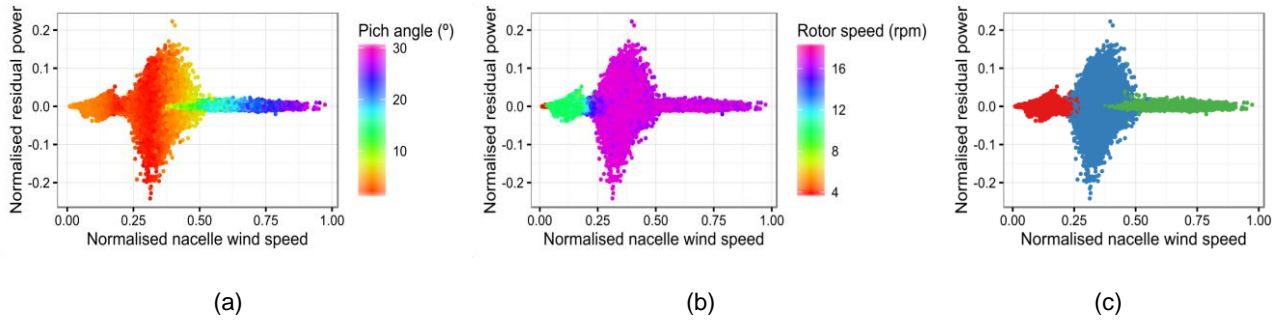


Figure 5. Normalised residuals varying according to pitch angle (a) and rotor speed (b), and data partitioning for mean and standard deviation calculation (c).

3. MODELLING ACCURACY RESULTS

The six suggested methods were applied to dataset WTH – 1 (see table 1). Modelling accuracy was evaluated through two performance metrics: mean absolute error (MAE) and the root mean squared error (RMSE). Results are summarised in table 2.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

Table 3: Results for modelling accuracy

| Model | MAE | RMSE |
|---------------------------------------|----------------|----------------|
| M1 – Method of bins | 0.02733 | 0.04876 |
| M2 – Generalised Linear Models | 0.01882 | 0.03651 |
| <i>Region 1</i> | <i>0.00089</i> | <i>0.00145</i> |
| <i>Region 2</i> | <i>0.00863</i> | <i>0.01164</i> |
| <i>Region 3</i> | <i>0.01715</i> | <i>0.02157</i> |
| <i>Region 4</i> | <i>0.06474</i> | <i>0.08210</i> |
| <i>Region 5</i> | <i>0.00432</i> | <i>0.00576</i> |
| M3 – k-Nearest Neighbours | 0.01868 | 0.03763 |
| M4 – Regression Trees | 0.05344 | 0.07764 |
| M5 – Random Forest | 0.00955 | 0.01951 |
| M6 – Support Vector Regression | 0.02257 | 0.03948 |

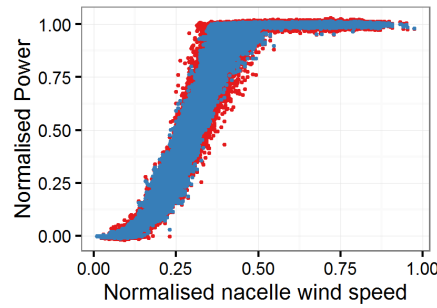


Figure 6. Real (red) and predicted power by M5 (blue)

4. CASE STUDY: PERFORMANCE MONITORING OF A FAULTY WIND TURBINE

Since M5 and M3 were found to be the more accurate, they were used for monitoring purposes. Models were trained with datasets WTH – 1 and WTF – 3, and then power residuals were monitored during the period from 01/02/2015 to 01/04/2015, using datasets WTH – 2 and WTF – 4. Residuals detected outside normal performance thresholds were flagged as events of underperformance. As a result, the number of detected events in each turbine from 01/02/2015 to 21/03/2015 (date where the WT4 was affected by a gearbox failure) are shown in figure 7.

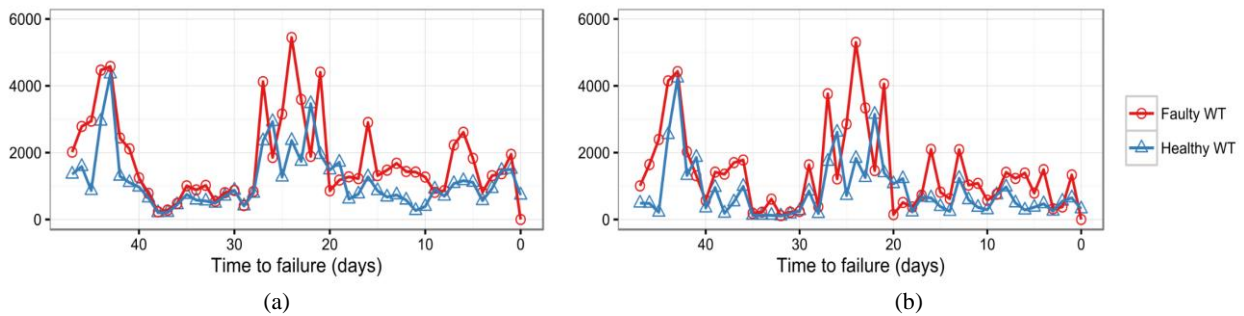


Figure 7. Number of underperformance events detected in each turbine before the failure. Power was predicted by M5 (a) and by M3 (b)

The number of events detected is greatly superior in WTF, especially from 30 days before the failure. As mentioned in section 2, categorised alarm logs from the SCADA were used to ascertain the reason for the underperformance detected. As a result, the breakdown of the detected events in both turbines is summarised in table 4. The number of detected events are presented as a percentage of observations.

Table 4: Events of underperformance detected for WTH and WTF categorised through the alarm logs

| Affected component / Cause | M5 | | M3 | |
|----------------------------|------------|-----------|------------|-----------|
| | Healthy WT | Faulty WT | Healthy WT | Faulty WT |
| Normal operation | 13.15% | 16.60% | 7.29% | 11.15% |
| Weather related events | 5.25% | 3.46% | 5.25% | 3.46% |
| Others | 3.95% | 13.07% | 3.55% | 12.08% |
| Yaw system | 0.28% | 0.90% | 0.38% | 0.90% |
| Gearbox assembly | 0.00% | 0.40% | 0.00% | 0.40% |
| TOTAL | 22.64% | 34.43% | 16.47% | 27.56% |

Both models detected similar number of abnormal performance during recorded issues, being in any case superior in WTF. However, both detected many underperformance events while turbines were supposed to

have a normal operation; this could be understood as false positives. The greatest difference between the two turbines is that the faulty turbine registered alarms related to the gearbox, while the healthy did not. This, together with the difference in number of total events detected could be considered as an indicator of the developing gearbox failure, proving therefore the detection capabilities of the proposed methodology. Finally, only critical components identified in [31] were highlighted; it would be necessary to consider a more detailed breakdown in order to understand which components were malfunctioning marked here as “Others”.

5. CONCLUSIONS

This paper presents a methodology where high-resolution SCADA data are used together with SCADA logs and suitably documented failure information for WT fault detection by monitoring its performance. The performance of a healthy turbine and a turbine affected by a gearbox failure are compared in a case study, proving the fault detection capabilities of the suggested approach.

Modelling results suggest that the use of high-resolution data improves accuracy. Nevertheless, further research should be performed to confirm this. Similar to previous contributions, non-parametric models were found to show better accuracy than parametric techniques. Indeed, the relation between wind speed and power is not governed by a unique mapping throughout the different operational regimes. As a result, accuracy in modelling benefits from a multivariate approach, that better accounts for power variability and permits a more conscious data partitioning. Nevertheless, multivariate non-parametric approach relies on including the whole gamut of data for the different predictors, and this may be difficult for some variables as, for example, the pitch angle; due to the pre-processing it is very likely that data registered during stall regime would be filtered out from the training dataset.

Finally, the assessment of uncertainty used for building the normal performance thresholds remains the strongest drawback of the suggested methodology, showing a high number of false positives in the case study presented. Further research should focus on improving uncertainty assessment for non-parametric models. Interesting approaches are Gaussian process and copula modelling for power curve modelling, since they provide predictions together with confidence levels.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642108 (AWESOME project - <http://awesome-h2020.eu/>). The authors would like to thank Enel Green Power for their cooperation and supply of the datasets, especially Giannis Loizos, Elena Lázaro and Jorge Asensio.

REFERENCES

- [1] European Wind Energy Association, “Wind in power,” no. February, pp. 1–12, 2017.
- [2] J. Carroll, A. McDonald, I. Dinwoodie, D. McMillan, M. Revie, and I. Lazakis, “Availability, operation and maintenance costs of offshore wind turbines with different drive train configurations,” *Wind Energy*, vol. 20, no. 2, pp. 361–378, Feb. 2017.
- [3] W. Yang, R. Court, and J. Jiang, “Wind turbine condition monitoring by the approach of SCADA data analysis,” *Renew. Energy*, vol. 53, pp. 365–376, 2013.
- [4] A. Albers, T. Jakobi, R. Rohden, and J. Stoltenjohannes, “Influence of meteorological variables on measured wind turbine power curves,” in *European Wind Energy Conference and Exhibition 2007, EWEC 2007*, 2007, vol. 3, pp. 1730–1736.
- [5] M. Dörenkämper, J. Tambke, G. Steinfeld, D. Heinemann, and M. Kühn, “Atmospheric Impacts on Power Curves of Multi-Megawatt Offshore Wind Turbines,” *J. Phys. Conf. Ser.*, vol. 555, p. 012029, 2014.
- [6] O. Uluyol, G. Parthasarathy, W. Foslien, and K. Kim, “Power Curve Analytic for Wind Turbine Performance Monitoring and Prognostics,” in *Annual Conference of the Prognostics and Health Management Society*, 2011, no. August, pp. 1–8.
- [7] S. Shokrzadeh, M. Jafari Jozani, and E. Bibeau, “Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods,” *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1262–1269, Oct. 2014.
- [8] A. Kusiak, H. Zheng, and Z. Song, “On-line monitoring of power curves,” *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, Jun.

- 2009.
- [9] M. Lydia, A. I. Selvakumar, S. S. Kumar, and G. E. P. Kumar, "Advanced Algorithms for Wind Turbine Power Curve Modeling," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 827–835, Jul. 2013.
 - [10] R. J. de Andrade Vieira, M. A. Sanz-Bobi, and S. Kato, "Wind turbine condition assessment based on changes observed in its power curve," in *2013 International Conference on Renewable Energy Research and Applications (ICRERA)*, 2013, no. October, pp. 31–36.
 - [11] E. Taslimi-Renani, M. Modiri-Delshad, M. F. M. Elias, and N. A. Rahim, "Development of an enhanced parametric model for wind turbine power curve," *Appl. Energy*, vol. 177, pp. 544–552, 2016.
 - [12] M. Schlechtingen, I. F. Santos, and S. Achiche, "Using Data-Mining Approaches for Wind Turbine Power Curve Monitoring: A Comparative Study," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 671–679, Jul. 2013.
 - [13] G. A. Skrimpas, K. Kleani, N. Mijatovic, C. W. Sweeney, B. B. Jensen, and J. Holboell, "Detection of icing on wind turbine blades by means of vibration and power curve analysis," *Wind Energy*, vol. 17, no. April 2013, p. n/a–n/a, 2015.
 - [14] E. Lapira, D. Brisset, H. Davari Ardakani, D. Siegel, and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach," *Renew. Energy*, vol. 45, no. SEPTEMBER, pp. 86–95, Sep. 2012.
 - [15] O. Uluyol and G. Parthasarathy, "Multi-turbine Associative Model for Wind Turbine Performance Monitoring," 2012.
 - [16] D. Karlsson, "Wind Turbine Performance Monitoring using Artificial Neural Networks With a Multi-Dimensional Data Filtering Approach," Chalmers University of Technology, 2014.
 - [17] E. Papatheou, N. Dervilis, a. E. Maguire, I. Antoniadou, and K. Worden, "A performance monitoring approach for the novel Lillgrund offshore wind farm," *IEEE Trans. Ind. Electron.*, vol. 0046, no. c, pp. 1–1, 2015.
 - [18] F. Pelletier, C. Masson, and A. Tahan, "Wind turbine power curve modelling using artificial neural network," *Renew. Energy*, vol. 89, pp. 207–214, 2016.
 - [19] M. Lydia, S. S. Kumar, a. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 452–460, 2014.
 - [20] G. A. Skrimpas, C. W. Sweeney, K. S. Marhadi, B. B. Jensen, N. Mijatovic, and J. Holboll, "Employment of Kernel Methods on Wind Turbine Power Performance Assessment," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 698–706, Jul. 2015.
 - [21] T. Ouyang, A. Kusiak, and Y. He, "Modeling wind-turbine power curve: A data partitioning and mining approach," *Renew. Energy*, vol. 102, pp. 1–8, 2017.
 - [22] A. Kusiak, H. Zheng, and Z. Song, "Models for monitoring wind farm power," *Renew. Energy*, vol. 34, no. 3, pp. 583–590, 2009.
 - [23] A. Marvuglia and A. Messineo, "Learning a Wind Farm Power Curve with a Data-Driven Approach," in *World Renewables Energy Congress 2011*, 2011, pp. 4217–4224.
 - [24] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Appl. Energy*, vol. 98, pp. 574–583, 2012.
 - [25] S. Butler, J. Ringwood, and F. O'Connor, "Exploiting SCADA system data for wind turbine performance monitoring," *Conf. Control Fault-Tolerant Syst. SysTol*, pp. 389–394, 2013.
 - [26] Jin Zhou, Peng Guo, and Xue-Ru Wang, "Modeling of wind turbine power curve based on Gaussian process," in *2014 International Conference on Machine Learning and Cybernetics*, 2014, pp. 71–76.
 - [27] Y. Wang, D. G. Infield, B. Stephen, and S. J. Galloway, "Copula-based model for wind turbine power curve outlier rejection," *Wind Energy*, vol. 17, no. 11, pp. 1677–1688, Nov. 2014.
 - [28] F. Castellani, A. Garinei, L. Terzi, D. Astolfi, M. Moretti, and A. Lombardi, "A new data mining approach for power performance verification of an on-shore wind farm," *Diagnostyka*, vol. 14, no. 4, pp. 35–42, 2013.
 - [29] O. Janssens, N. Noppe, C. Devriendt, R. Van de Walle, and S. Van Hoecke, "Data-driven multivariate power curve modeling of offshore wind turbines," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 331–338, 2016.
 - [30] M. Reder, E. Gonzalez, and J. J. Melero, "Wind Turbine Failure Analysis - Targeting current problems in Failure Data Analysis," in *The Science of Making Torque from Wind*, 2016.
 - [31] E. Gonzalez, M. Reder, and J. J. Melero, "SCADA alarms processing for wind turbine component failure detection," *J. Phys. Conf. Ser.*, vol. 753, no. 7, p. 72019, 2016.
 - [32] International Eletrotechnical Comission, "IEC 61400-12-1 Wind turbines – Part 12-1: Power performance of electricity-producing wind turbines," 2011.
 - [33] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY: Springer New York, 2013.
 - [34] A. Kusiak, H. Zheng, and Z. Song, "Power optimization of wind turbines with data mining and evolutionary computation," *Renew. Energy*, vol. 35, no. 3, pp. 695–702, 2010.
 - [35] A. Clifton, L. Kilcher, J. K. Lundquist, and P. Fleming, "Using machine learning to predict wind turbine power output," *Environ. Res. Lett.*, vol. 8, no. 2, p. 024009, 2013.
 - [36] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. New York: Chapman and Hall, 1984.
 - [37] L. Breiman, *Random Forests*, vol. 45, no. 1. Springer, 2001.
 - [38] A. Kusiak and A. Verma, "A Data-Mining Approach to Monitoring Wind Turbines," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 150–157, Jan. 2012.

Estimation Method of Static Contact Wire Height by Using Contact Force on Commercial Line

Takayuki Usuda ¹, Mitsuru Ikeda ²

¹ Current Collection Laboratory, Railway Dynamics Division, Railway Technical Research Institute, Hikari-cho 2-8-38, Kokubunji-shi, Tokyo, Japan

² Railway Dynamics Division, Railway Technical Research Institute, Hikari-cho 2-8-38, Kokubunji-shi, Tokyo, Japan

ABSTRACT

The overhead catenary system (OCS) is one of the most important subsystems for electric railway. Due to absence of their redundancy, the accidents between the OCS and the pantograph result in long operational down-time. To prevent such accidents, a large amount of manpower is necessary for its maintenance work. However, labor-saving related to the maintenance works has become a very important subject in preparing the incoming of aging society in Japan.

Installation tolerance and maintenance criteria of the OCS (contact wire height, gradient, stagger, overlap configuration, and so on) are specified for the static condition. These values are checked by the maintenance personnel. Maintenance can therefore be made efficiently if it is possible to estimate these static geometries with the data obtained by an inspection vehicle. The current inspection vehicle however measures the pantograph height instead of the static contact wire height. The pantograph height is not identical to the static contact wire height since the pantograph height includes dynamic uplift of the contact wire.

In this paper, the authors propose the method to estimate the static contact wire height for OCS systems by means of the data measured by the inspection vehicle. The estimation results by this method on two kind of OCS types agree very well with the measured static height of contact wire.

Keywords: Diagnosis, Pantograph Contact Force, Static height of contact wire, Maintenance

Corresponding author: Mr. T. Usuda (usuda.takayuki.99@rtri.or.jp)

1. INTRODUCTION

OCS is one of the important equipment of an electric railway. Due to absence of redundancy, accidents result in long operational down-times. To prevent such accidents, a large amount of manpower is necessary for its maintenance work. At the same time, given the incoming of ageing society, Japan faces an urgent need to reduce labor-intensive maintenance. Recently a large amount of researches have been conducted in order to effectively use the data collected by the inspection vehicles to facilitate maintenance ^{[1] [2] [3]}. For reducing the maintenance cost the condition based maintenance methods have been also studied ^[4].

Installation tolerances and maintenance criteria for the OCS (i.e. contact wire height, gradient, deviation and overlap configuration, and so on) are specified for static geometry of the OCS. These indices are checked by the maintenance personnel. Maintenance would therefore be easier if it is possible to verify the static geometry from the inspection vehicles. This boosted the need for a method which can evaluate the static geometry of the OCS.

The current inspection vehicles however measure the pantograph height instead of the static contact wire height; the pantograph height is not identical to the static contact wire height since the pantograph height includes dynamic uplift of the contact wire. Stereo image processing techniques show promise for static geometry measurement, but it is not easy to be applied to on-board measurement at high speed.

The authors have been studying a method to estimate the static contact wire height from the data measured by inspection vehicles [5]. This paper proposes a method to estimate the static contact wire height of the OCS with the span lengths variation. In this method, as a pre-processing, the authors numerically obtain the dynamic behaviour of OCS when vibration forces being applied to the contact wire in several cases. From these simulation results the dynamic characteristics of the OCS are calculated. The static height of contact wire can be estimated from these characteristics, the pantograph contact force and the pantograph height. It also presents the results of verifications of the proposed method through the line tests results of two kind of OCS types.

2. ESTIMATION METHOD OF STATIC HEIGHT OF CONTACT WIRE

2.1. Transfer function method

To estimate the static contact wire height, this section proposes a transfer function method. The static contact wire height is able to be estimated by this method taking into account the fluctuation of the equivalent stiffness of the contact line in the hanger span interval. The simulated excitation test of the contact line is executed in advance. The static contact wire height $y'_0(x)$ is estimated from the contact force and the height of the first pantograph, i.e. $f_1(x)$ and $y_{pan1}(x)$, by using dynamic property of the contact line calculated from the simulated excitation test. The details of the proposed method are described below.

First, excitations of a contact wire all of whose static height at hanger points are identical are simulated for 10sec by changing the excitation point. In this procedure the excitation points are set on the contact wire with an interval of 1 meter in the 200m section, and OCS excitation simulation is carried out with a time step of $\Delta t = 0.1\text{msec}$. The excitation force is a sum of DC wave of 1N and random wave with variation from -100N to 100N. The DC component in the excitation force is necessary to reflect the effect of the pantograph uplift force. By using results of the excitation simulation, the transfer function $H_{ij}(\omega)$ from excitation force acting on the point i to contact wire height at the observation point j can be evaluated as below:

$$H_{ij}(\omega) = \frac{Y_j(\omega)}{F_i(\omega)} \quad (2)$$

$Y_j(\omega)$: Fourier transformation of the contact wire height at observation point j

$F_i(\omega)$: Fourier transformation of the excitation force at excitation point i

The impulse response function $h_{ij}(t)$ between the excitation point i and the contact wire height observation point j was calculated, by computing the inverse Fourier transform of the transfer function $H_{ij}(\omega)$. When the pantograph is running at velocity v , the predicted value of the contact wire uplift $y'_{up1}(x)$ can be calculated based on convolution integral of the first pantograph contact force $f_1(i)$ at point i and the impulse response function as shown below:

$$y'_{up1}(x) = \sum_{i=1}^m h_{ij}(x - i \cdot \Delta x / v) f_1(i \cdot \Delta x / v). \quad (3)$$

Because the pantograph is moving, the impulse response functions evaluated at each excitation point are needed. The contact force sampling interval Δx was set to be 1m and the sampling rate was adjusted depending on the train speed. To simplify notation, equation (3) is rewritten with an impulse response matrix \mathbf{h} as shown below:

$$\begin{bmatrix} y'_{up1}(x_m) & \cdots & y'_{up1}(x_{m+50}) \end{bmatrix}^T = \mathbf{h} \begin{bmatrix} f_1(x_1) & \cdots & f_1(x_i) & \cdots & f_1(x_{m+50}) \end{bmatrix}^T. \quad (4)$$

By preparing an impulse response matrix for each train speed in advance, the uplift of the contact wire can be estimated in the estimation area from equation (4), based on the measured value of the pantograph contact force of the first pantograph. Using the estimated value of the contact wire uplift $y_{up1}'(x)$ and the pantograph height $y_{pan1}(x)$, it is possible to estimate the static contact wire height $y_0'(x)$ with the following equation.

$$y_0'(x) = y_{pan1}(x) - y_{up1}'(x) \quad (5)$$

2.2. Verification in the line test result of compound catenary

The static contact wire height on a commercial line was estimated by the above mentioned method. In this procedure, the dynamical simulation of excitation of the compound catenary is executed in this section. The excitation points are set on the contact wire with an interval of 1 meter in the 200m section as shown in figure 1. The tensions of the respective wires of the contact line used in this simulation are shown in table 1. The sampling frequency of the simulation was set to be 10 kHz and the sampling rate was re-sampled at a lowered frequency. Because the pantograph moving at 270km/h in this section, the sampling rate was set to 75 Hz which corresponds to 1m on spatial resolution. Based on this resampled data, the transfer function was calculated. A graphic representation of the calculated impulse response matrix of equation (4) is shown in figure 2. In this figure the origin of the XY axis is the support point which is the start point of estimation area of contact wire height.

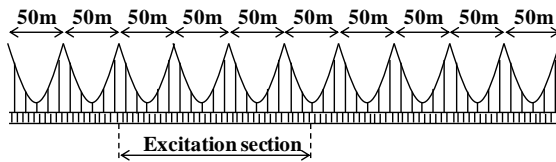


Figure 1. Calculating condition of transfer functions

Table 1: Tensile forces of contact line

(High speed lines)

| | Tension (kN) |
|----------------|--------------|
| Messenger wire | 19.6 |
| Auxiliary wire | 14.7 |
| Contact wire | 19.6 |

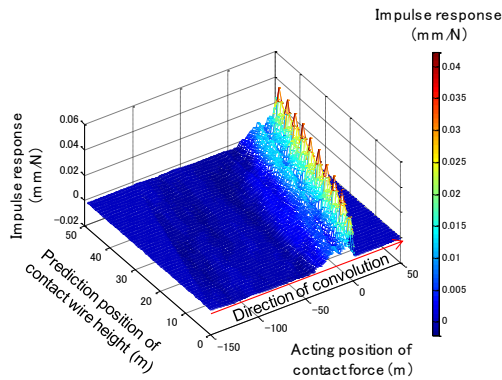


Figure 2. Calculation result of impulse response matrix.
(v=270km/h)

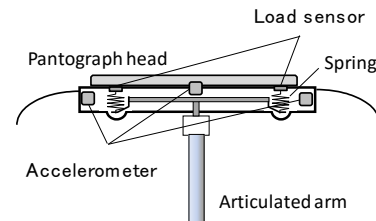


Figure 3.. Sensor setup for contact force measurement.

The contact force of the pantograph and the pantograph height are measured at 270 km/h on actual high speed train. The contact force of the pantograph running was also measured by the accelerometers and the load sensors embedded in the pantograph head. The arrangement of sensors in the pantograph head is shown in figure 3. The inertial force of the pantograph head was measured by the accelerometers and the force applied on the pantograph head by the articulated frame was measured by the load sensors, respectively. The contact force was derived from them [2]. The contact force measured using this method was sufficiently accurate for frequencies below 60 Hz. The pantograph height on the other hand was measured by sensing the angle of the lower frame of the pantograph. Based on these measurement results, the static contact wire height was estimated by equation (5) in about 1000 m long section. For verification of this estimation the static contact wire height was also measured by temporary measuring instrument set on a top of catenary maintenance vehicle at very low speed.

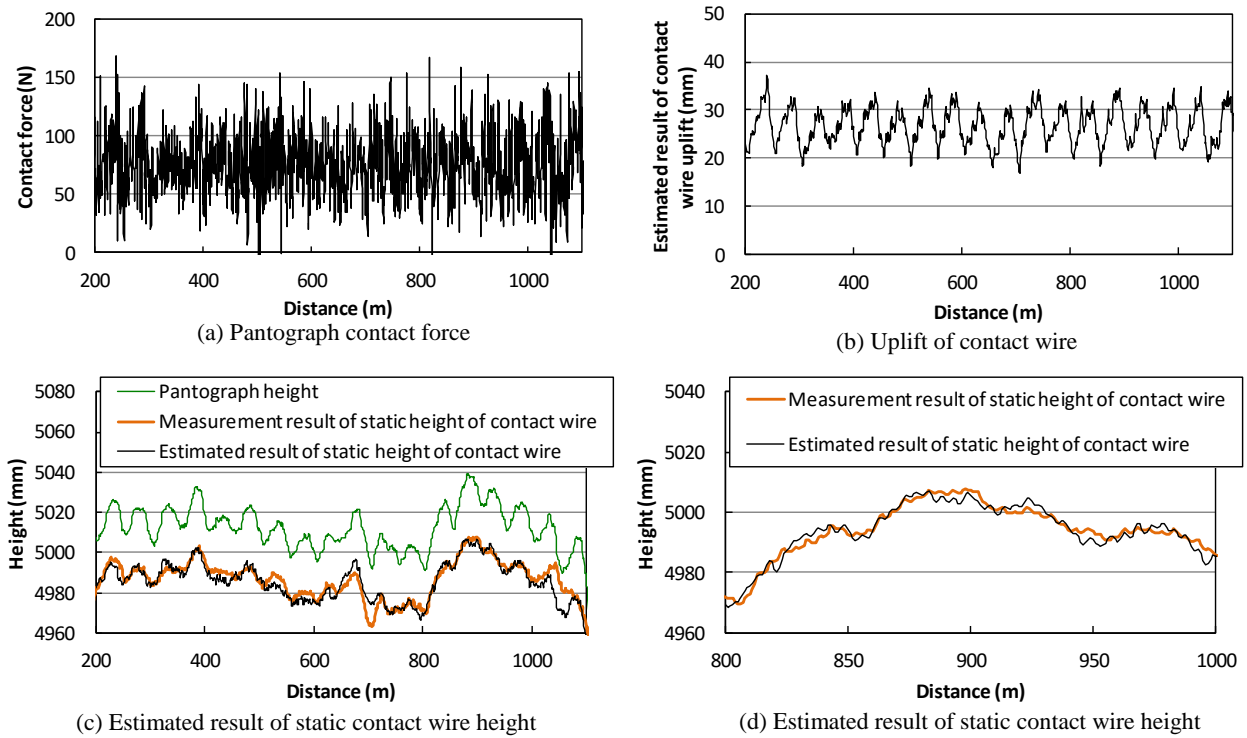


Figure 4. Estimated result of static contact wire height by using transfer function method.

The comparison between the estimated static contact wire height by the transfer function method and the measured one by the maintenance vehicle is shown in figure 4. The line test data were measured for a train running at 270 km/h. Figures 4(a)-(c) show the contact force, the estimated results of the uplift of contact wire and the estimated results of the static contact wire height, respectively. It should be noted that the distance points of every 50m from 200m to 1100m correspond to the support points of OCS. Figure 4(d) is the extended version of figure 4(c). From this result, the estimated static contact wire height sufficiently agrees with the measured height. Figure 4(d) shows that the proposed method can estimate the static contact wire height within an accuracy of about 4 mm for most section. The reason why the estimation result is including some err at the distance position 700m may be considered as effect of vertical vibration of train.

2.3. Verification in the line test result of simple catenary with span lengths variation

In the previous section, we apply the proposed method to the compound catenary. Because the compound catenary have the higher evenness of OCS vertical stiffness than simply catenary in general, the contact wire static height of compound catenary can be estimated with relatively little effort. In this section, we apply the proposed method to measurement results in line equipped with simple catenary. And the versatility of the proposed method to estimate the static contact wire height with span lengths variation are verified. As an example, the case where all the span lengths of the OCS are 50m except for one span, as shown in figure 5, is tested. The exceptional span length is 45m.

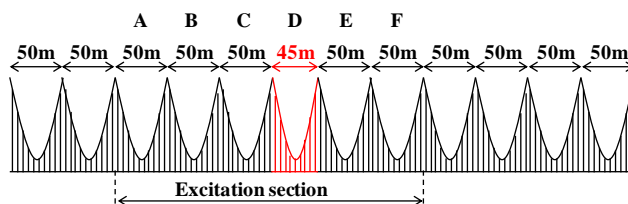


Figure 5. Calculating condition of transfer functions

Table 2: Tensile forces of contact line

(Conventional lines)

| | Line density (kg/m) | Tension (kN) |
|----------------|------------------------|-----------------|
| Messenger wire | 0.70 | 19.6 |
| Contact wire | 1.51 | 14.7 |

The specifications of the respective wires of the contact line used in this simulation are shown in table 2. The excitations of a contact wire are simulated in the same way as that indicated in the section 2.1. From

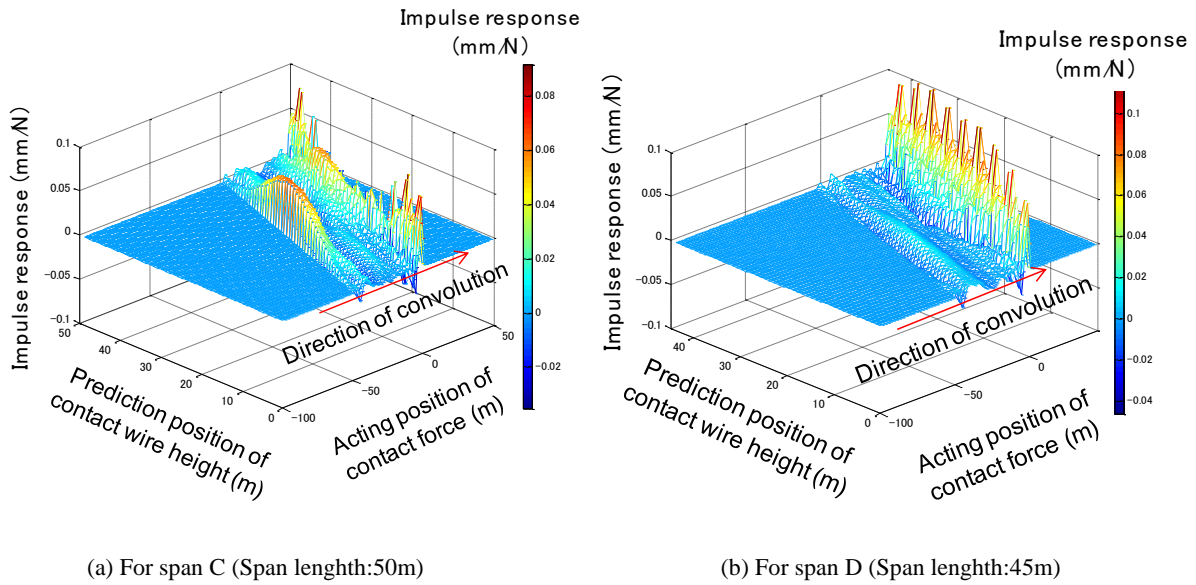


Figure 6. Calculated result of impulse response matrix

these excitation data, impulse response matrixes to estimate the uplift of the span from C to F were calculated. A graphic representation of the calculated impulse response matrix is shown in figure 6. Figures 6(a)-(b) show impulse response matrixes to estimate the uplift of the span C and the span D respectively.

The static contact wire heights with the span lengths variation were estimated using the above method. The static contact wire height for this commercial line was actually measured by a catenary maintenance vehicle as well. The contact force of the pantograph running at 127 km/h was also measured by fixing the accelerometers and the load sensors on pantograph head as indicated in the section 2.2. The pantograph height was measured by laser displacement sensor. Based on these measurement results, the static contact wire height was estimated with impulse response matrixes. The comparison between the static contact wire height estimated with the transfer function method and the measured one by separate work on the maintenance vehicle is shown in figure 7. The length of span from the distance position 400m to the distance position 445m is 45m and all of the other span lengths are 50m.

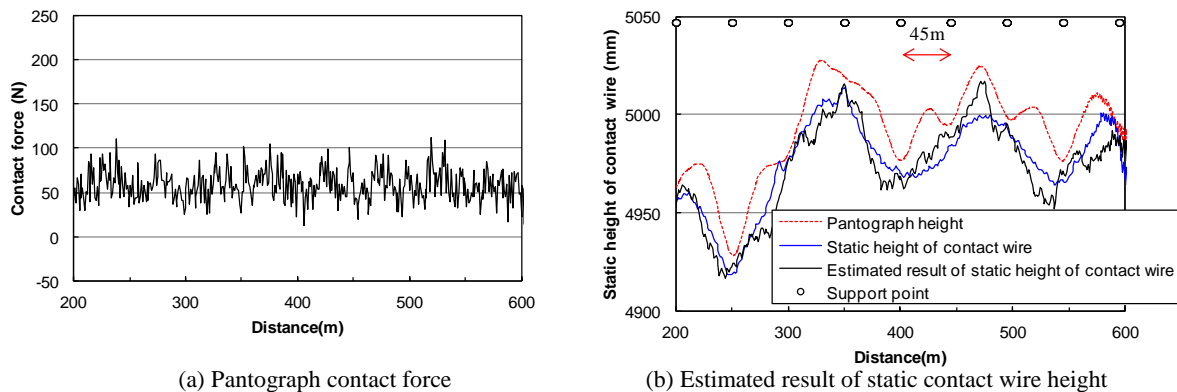


Figure 7. Estimated result of static contact wire height by using transfer function method.

Figures 7(a)-(b) show the contact force and the estimated results of the static contact wire height, respectively. From the result of figure 7(b), it was found that the estimated static contact wire height sufficiently agrees with the measured height within an accuracy of about 20 mm.

3. CONCLUSIONS

This paper proposes a method to estimate the static contact wire height from the measured data collected by the inspection cars. This method is the transfer function method: the contact wire uplift due to the passing pantograph is estimated by the contact force and the impulse response matrix calculated from the pre-processing simulated excitation tests of the overhead contact line. Using the pantograph height and the estimated value of the contact wire uplift, it is possible to estimate the static contact wire height. The conclusions of this paper are as follows:

- (1) The transfer function method can estimate the static contact wire height of compound catenary from the pantograph height and the pantograph contact force within an accuracy of about 4 mm for most section.
- (2) The static contact wire height of the simple catenary can be estimated from the line test with span lengths variation.

REFERENCES

- [1] T. USUDA, "Estimation of Wear and Strain of Contact Wire Using Contact Force of Pantograph", Quarterly Report of RTRI, Vol.48, No. 3, pp.170-175, 2007.
- [2] M. IKEDA, "The Contact Force between Pantograph and Contact Wire - An Estimation Method Using the Inversion Technique -", Quarterly Report of RTRI, Vol.45, No. 2, pp.80-85, 2004.
- [3] T. KOYAMA, et al., "Measurement of the Contact Force of the Pantograph by Image Processing Technology", Quarterly Report of RTRI, Vol.55, No. 2, pp.73-78, 2014.
- [4] M. CORFIATI, et al., "Evolution lines in the maintenance of rolling stock", *Ingegneria Ferroviaria*, pp.751-772, Vol. 66, No.9, 2011.
- [5] T. USUDA, "Estimation Method of Static Height of Contact Wire by Using Pantograph Contact Force", Quarterly Report of RTRI, Vol.56, No. 3, pp.175-180, 2015.

Feature Selection for Remaining Useful Life Prediction of Spur-Gears

K. López de Calle¹, S. Ferreiro¹, E. Konde¹, I. Bravo¹, A. Arnaiz¹

¹ Intelligent Information Systems, IK4 - TEKNIKER, Eibar 20600, Spain

ABSTRACT

Gears are common components in all type of equipment and systems and play a key-role in crucial ones such as gearboxes. Real understanding of failure modes is important in order to maintain them correctly, because any defect can result in machine downtime. Predictive maintenance is getting over corrective and time based preventive maintenance, where two different streams had been distinguished in its development: Model-driven and Data-driven approaches. But, for them to be possible to work properly, the election of reliable signals and descriptors is crucial. Conventionally used condition monitoring techniques such as vibrations or even acoustic emissions are known to be good for prediction. Nevertheless, this work focuses on the selection of the best signals and descriptors among different signals originated in a FZG testing bench for spur-gear pitting analysis. Data is differently processed, treated and studied with multiple data-mining methods and then the results are compared using for this purpose commonly used machine learning algorithms.

Keywords: Image processing and machine learning for maintenance; Monitoring, diagnosis, prognosis and health management; Sensors technology and damage detection.

Corresponding author: Kerman L. (email: kerman.lopezdecalle@tekniker.es)

1. INTRODUCTION

Spur-gears are basic parts of the machinery as known nowadays. They can be found in a wide variety of machines with completely disparate functions, such as wind turbines, cars, machine tools or gearboxes. Such fundamental component of a machine has to be correctly maintained to avoid breakdowns or undesired behaviours. This work focuses on pitting which is one of the mayor causes of wear on gears. Measuring that wear is not as straightforward as it may seem, as the machine has to be stopped and the gears dismantled and analysed to measure that wear. Nevertheless, this method is expensive and time consuming, that is why much research has been carried out with the aim of developing a more efficient method for the assessment of remaining useful life (RUL). Most of those methods are based in condition monitoring, which consists in monitoring signals released by the machine while it is working and detecting the evolution of anomalies in the signals.

It might be noticed that most of the research about condition monitoring has focused on vibration analysis due the reliability of this kind of signals [1]. Other signals such as acoustic emissions [2], [3] or current [4], have also proved to be useful for gear condition monitoring. However, accurate measurements of pitting and early detection in gears is hardly found in literature. This is because the difficulty all those methods face of assessing the amount of pitting in real time. As the only way of assessing pitting is dismantling the gears and inspecting them using a microscope. Signals provide a measure of the deviation of the machine from the regular operation. From this deviation, the state of pitting is inferred.

To avoid undesired damage to happen, assessing pitting state of the gears in early stages is needed, so that preventive measures can be taken before critical damage of components appears. For this purpose, the election of adequate signals and descriptors is crucial. This work aims to determine which combination of signals and predictors are most appropriate for remaining useful life prediction of certain machinery, as well as which machine learning prediction algorithms perform better in this case.

Among the various potential hazards spur gears might suffer, pitting is cause of nearly 60% of gear failures [5]. Pitting is a type of surface fatigue originated when material's endurance limits are exceeded. This

failure depends on surface contact stress and number of stress cycles. At early stages, small pits appear in localized, over-stressed areas. Most of the times this phenomenon arises due to the inappropriate alignment across the full face width of the gear mesh. If not handled properly, pitting initiation (a.k.a. micro-pitting) can lead to considerably large pitted surfaces with larger diameters. This larger pitting (macro-pitting) will continue until gear tooth profile is completely destroyed, causing extremely rough operation and considerable noise, which often ends up with premature tooth breakage failure [6].

With the purpose of detecting best predictors for micro-pitting assessment, several tools commonly used in data-mining have been employed and the results analysed. Those techniques and results are explained in more detail in next chapters.

2. METHODS

2.1. Experimental setup

The data analysed in this work was generated in a FZG test rig. This kind of work benches are standardized and well known for spur gear testing. They follow a back-to-back or power circulation configuration and allow to perform different standardized experimental test for gears, such as gear scuffing tests, gear micro-pitting tests, gear pitting tests etc. Diagram of a FZG rig is shown in figure 1.

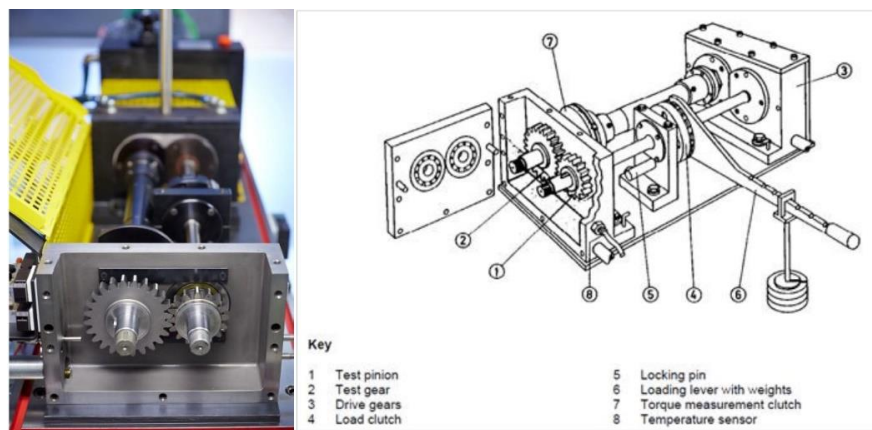


Figure 1: FZG test rig diagram.

Regarding the spur gears used for the data generation, ten couples of wheel/pinion gears were used for the tests. Wheels had 42 teeth while the pinions had 28, regarding the pressure angle, some of the couples had 20 degrees while some others had 22.5.

2.2. Extracted features

The dataset obtained from the tests consisted in the signals produced by 10 different spur-gear couples during the operation. In addition, some micro-pitting measures were taken every 24 hours after the first 72 hours of operation. Data acquisition frequency was set to be 10.420 Hz, during 56 seconds every 15 minutes. As explained before, after 24 hours the test rig was stopped to measure micro-pitting and it was launched again. Due to the huge amount of data to be processed, for each of the files with 56 seconds a set of statistical descriptors was extracted from each signal. Filtering was done using 1 second windows and averaging obtained values, as more complex filtering techniques are beyond the object of study of this work. It should be noted that this work has been done from pre-existing data, which limited the ability to modify taken signals or acquisition criteria.

Besides common statistical predictors, specific descriptors were also extracted for the case of vibration and microphone signals, as the bibliography clearly supports the achievements reached with the analysis of

side-band of gear mesh frequencies of vibrations [1] and the energy index of the microphone [3]. For the extraction of side bands Fourier Transforms were used. Once the 56 second signals were converted to frequency domain, the maximum value and the root mean square (RMS) values of ± 10 Hz windows and the RMS value for ± 30 Hz windows were taken. This was done for the first 5 harmonics.

In the case of the microphone, in order to force the energy index (EI) to fit the amount of data taken in the other descriptors, first the signal with the 56s was divided in different parts using for that purpose the position of the encoder so that each part was equal to a revolution; then the RMS of each part was computed as well as the RMS of the whole signal. Last, 4 different power values were used for EI calculation, 1, 3, 5 and 10.

$$EI = \left(\frac{RMS_{segment}}{RMS_{total}} \right)^k$$

Table 1: Signals and respective descriptors.

*Pressure angle: Introduced as a factor variable with two values in relation to the shape of the gear (20/22.5).

**H: 10Hz and 30Hz windows have been taken for the first five harmonics.

*** Energy Index: Different values of power have been used (1,3,5,10).

| Descriptors Signals | RMS | Peak Value | Crest Factor | Average | Standard Deviation | kurtosis | Skewness | Clearance | Impulse | Shape | Median | Max | Min | H** | Max | H** | 10Hz_RMS | H** | 30Hz_RMS | Energy Index *** |
|------------------------|-----|------------|--------------|---------|--------------------|----------|----------|-----------|---------|-------|--------|-----|-----|-----|-----|-----|----------|-----|----------|------------------|
| Current Intensity | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | |
| Acelerometer X axis | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | |
| Acelerometer Y axis | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | |
| Acelerometer Z axis | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | |
| Microphone | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | x |
| Encoder rpm | x | x | x | x | x | x | x | x | x | x | x | x | x | | | | | | | |
| Pressure Angle* | | | | | | | | | | | | | | | | | | | | |

Finally, the data set contained a total of 128 variables as seen in the table, with the additional wheel and pinion degradation variable.

2.3. Data-mining techniques

2.3.1. Complexity of the problem

As explained before, the complexity of this work lies in the difficulty of assessing the real value of pitting in real time. Data was taken over the whole process, however, the only labelled values were the ones at the end of each test. As result, only 39 out of more than 5 thousand instances of the dataset had label. To overcome this problem, linearity was assumed as hypothesis in the evolution of pitting.

Another difficulty to be faced was the exclusion of temporality as one of the variables. This decision was adopted considering the aim of the work was to assess the state of the gears aloof the working time of the gear (classical RUL), so that a real condition based maintenance could be done. Undoubtedly this greatly increased complexity of the task, but would increase the chance to apply it in real life cases.

Last complication to mention is the fact that there was no a single class, but a double class, as the degradation was not equal in both wheel and pinion and did not evolve in the same way, even if they were highly correlated. This issue was sorted out by using just one of the gears for predictions, as both degradations (from wheel and pinion) are correlated, as it will be explained later.

2.3.2. Data exploration/cleansing

An initial exploration of the data was done to detect any possible outliers or misleading features. It was found that even though the data acquisition started after a running-in period of time, the machine was not

yet working in stable conditions. This was observed after a Principal Component Analysis was done and outliers were found before the stable working conditions were reached. Once the temperature of the lubricant is stabilized, the values of the signals take also steady values, so the data set was cleaned using the temperature of as indicator.

Another interesting observation was the fact that the highest correlation found among predictors and class was given by the degradation of the pinion with the degradation of the wheel. When one of the degradations was used for foreseeing the degradation of the other the results improved significantly. However, it must be noticed that this data was created assuming the hypothesis of linearity and it was not supposed to be used for prediction, but the relation among wheel and pinion degradation might be interesting to analyse in future works.

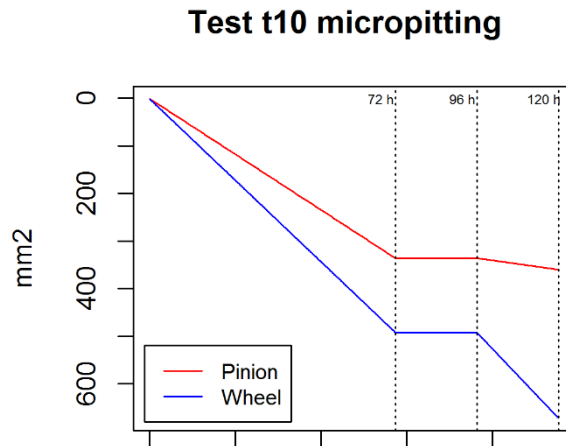


Figure 2: Evolution of micro-pitted surface in gear/pinion couple number 10.

2.3.3. Feature subsets selection

One of the relevant parts of prediction in data mining process to extract knowledge from data, is the feature subset selection. The election of irrelevant, redundant or noisy data hinders the prediction for machine learning algorithms [7]. Furthermore, the feature subset selection improves the understanding of the data, reduces storing requirements and shortens training times [8].

Amongst the automated variable selection algorithms two main groups must be distinguished, filter method and wrapper method algorithms. As there is no unique solution for subset selection, this work compares the data subsets of variables obtained with different feature selection algorithms (2 filter methods and 2 wrapper methods) and evaluates the feature subsets by means of regression algorithms to decide which subset performs better in this particular case.

2.3.4. Filter methods:

Filter methods evaluate the performance of the variables without taking the prediction model into account. Some criteria are used to determine whether the variable is relevant for the forecasting or not. Filter methods are computationally efficient and robust against overfitting. However, the risk of not identifying redundant data or ignoring important interactions among variables may increase with filter methods. The following filter methods were used for the feature subset selection:

2.3.5. Correlation Feature Selection (CFS):

CFS main assumption is that good features are correlated with the class, while they are not correlated with the rest of the features. From this premise, different feature subsets are evaluated and ranked with a 'merit' value and the best subset is chosen. In this merit value, number of features, average class/feature correlation and average feature/feature correlation (redundancy) are considered [7].

2.3.6. Regression Relief (RRelief):

RRelief is a variation of Relief heuristic method for attribute estimation, which focuses in estimating the quality of continuous class attributes. Original Relief algorithm (for discrete class) estimates the attributes' value in relation to how well their values distinguish between close instances. Given a random number of variables (R), it searches for nearest neighbours: One with the same class and other with different class, called nearest hit (H) and nearest miss (M) respectively. Then, a quality estimator (W) is updated for all the attributes (A), in relation to the values R , M and H . This process is repeated as many times as the user desires (m). The update of the quality estimator differs in the cases of continuous or discrete variables being the goal the same, to measure the distance among attributes [9].

2.3.7. Wrapper methods:

In contrast to filter methods, wrapper methods evaluate the performance of the attributes using for this purpose an arbitrarily selected classification algorithm. On the one hand, this eases the search of interactions among variables and the method decides the subset in relation to the performance of the estimation. On the other hand, this kind of methods have a main drawback: The use of an algorithm for attribute selection makes it easier to overfit the selection. Which means that the method could choose a feature subset optimized for the search algorithm without guaranteeing this subset will optimize other regression algorithms' performance. Furthermore, the need of training prediction algorithms several times makes these methods computationally slow.

It was decided to use two different wrapper method algorithms applied to random forest algorithm.

2.3.8. Backwards selection (Back):

Backwards selection starts from a dataset with all the variables, it creates a model with the chosen algorithm and ranks the variables regarding the importance they have for the development of the model. Variables with lowest importance values are discarded and the model is trained again to see whether the prediction improves or not.

2.3.9. Genetic Algorithms (GA):

Genetic algorithms are well known heuristic search algorithms. With an initial population (a set of variables) they measure the quality of the members (the accuracy of the model) and rank them. Then new members (combinations of variables) are created by crossing and mutating individuals. Individuals are ranked again and a selection is done keeping just those with the highest quality values. This process is repeated over several generations or times the user decides. The best subset is the one with the highest quality value at the end of the required generations.

2.3.10. Validation

The four subsets of variables obtained from features selection methods were evaluated by using different regression algorithms. In order to have statistically significant results, data was split into three independent datasets: one to analyse and obtain most relevant subset of features, other to train the algorithms and finally another one to test them. In this way, the statistical analysis about the goodness of the model was evaluated and it ensured the independency between data partitions (training and testing data). Due to the nature of the data, as it comes from different gears, instead of splitting the data in random subsets it was split according to the gear they belonged to. Data coming from first 4 gears was used for feature subset selection, another 4 were used for training models and the last 2 for testing.

As the behaviour of classification/regression algorithms is different due to their own fundamentals, the algorithms used for validation belonged to different algorithm families. there were selected dissimilar types

of them with diverse types of basis to test the performance of each of the previously selected feature subsets:

- Bayesian Ridge Regression: Probability based algorithm slightly modified for regression cases.
- KNN: Estimates class value averaging the one of the k nearest neighbours. The algorithm groups data in groups in such way that the elements of one group have high similarity between them but not with the elements of other groups.
- Decision Tree: Studies various variable splits to determine which split provides minimum variance, therefore increasing the probability of belonging to one class.
- Random forest: Combination of Decision Trees. With the added advantage of correcting overfitting
- Conditional forest: Variant of random forest that uses non-parametric test as splitting criteria to avoid overfitting.

3. RESULTS

Initially, feature selection methods and validation algorithms were used to predict the pitting from raw data. The results are shown in table 2. To have a right interpretation of the results it is needed to mention that Relief algorithm was used for future prediction considering 7 and 75 variables. These values were assumed taking into account the number of features selected by CFS, Backwards and Genetic algorithms.

Table 2: RMSE results obtained from initial data.

| | Initial | | | | | | Mean | Min |
|-----------|-----------|---------------------|---------------|--------|------------------|--------------------|-------|-------|
| | Variables | Bayesian Regression | Decision Tree | KNN | Random Forest | Conditional Forest | | |
| Raw | 128 | 146,54 | 158,42 | 96,16 | 94,32 | 111,24 | 121,3 | 94,3 |
| CFS | 1 | 67,99 | 130,37 | 156,30 | 82,80 | 68,73 | 105,8 | 68,0 |
| CFS | 5 | 90,94 | 56,39 | 128,21 | 82,80 | 46,31 | 80,9 | 46,3 |
| CFS | 6 | 138,05 | 56,39 | 127,99 | 103,15 | 45,31 | 94,2 | 45,3 |
| Relief | 7 | 158,74 | 171,07 | 184,09 | 170,68 | 62,40 | 149,4 | 62,4 |
| Relief | 75 | 240,10 | 158,27 | 124,79 | 125,38 | 104,00 | 150,5 | 104,0 |
| Backwards | 85 | 246,92 | 158,41 | 98,41 | 85,98 | 42,21 | 126,4 | 42,2 |
| Genetic | 70 | 106,43 | 158,42 | 73,99 | 112,28 | 46,42 | 99,5 | 46,4 |
| | Mean | 149,5 | 131,0 | 123,7 | 110,7 | 65,8 | | |
| | Min | 68,0 | 56,4 | 74,0 | 82,8 | 42,2 | | |

As expected, average values of Decision Tree, Random and Conditional Forests overcame the performance of Bayesian Regression or KNN algorithms. The main reason for this is that Backwards and Genetic algorithms already used Random Forest algorithm to choose variables. As Conditional Forest and Decision Tree belong to the same family of algorithms, their performance was also better. During the feature selection process, surprisingly, CFS algorithm chose a single variable, 'Accelerometer_Y_H1_Max', which was not compatible with Random Forest as it needed at least 2 different variables. It was decided to run the CFS algorithm again without that variable to see which was the second-best subset. After, two extra subsets were considered, one with the sum of variables of first and second subsets and another with just the second-best subset's variables.

Another interesting conclusion is that more is not always synonym of better, the lowest mean error was reached by a subset of 5 variables chosen by the CFS algorithm. However, the lowest error (RMSE of 42 points, Backwards selection with Conditional Forest for prediction) does not seem to be good enough, knowing that this real value ranges from 0 to 672 mm² in our samples.

In the beginning, algorithms were supposed to be tested just with different subsets taken from the raw data. However, the omission of the temporality and transience of signals increased the difficulty of finding patterns because each test started with slightly different values in certain period of time due to the noisy operation environment. To normalize data and, in certain way, to have some control over the temporality without adding time parameter explicitly, it was decided to use finger print concept. That is, the first value

of the signal was used for dividing the whole signal, so that the rest of the values became distances to that initial value.

Table 3: RMSE results obtained from normalized data.

| | Normalized | | | | | | Mean | Min |
|-----------|------------|---------------------|---------------|--------|---------------|--------------------|------|------|
| | Variables | Bayesian Regression | Decision Tree | KNN | Random Forest | Conditional Forest | | |
| Raw | 128 | 138,24 | 76,11 | 51,06 | 40,77 | 43,24 | 69,9 | 40,8 |
| CFS | 6 | 35,45 | 52,71 | 116,10 | 42,32 | 53,64 | 60,0 | 35,5 |
| Relief | 7 | 59,15 | 56,60 | 70,30 | 51,95 | 58,05 | 59,2 | 51,9 |
| Relief | 75 | 129,77 | 68,53 | 65,89 | 35,86 | 43,09 | 68,6 | 35,9 |
| Backwards | 50 | 93,24 | 76,11 | 43,32 | 47,48 | 43,08 | 60,6 | 43,1 |
| Genetic | 71 | 117,17 | 89,18 | 78,10 | 37,59 | 44,21 | 73,3 | 37,6 |
| | Mean | 95,5 | 69,9 | 70,8 | 42,7 | 47,6 | | |
| | Min | 35,5 | 52,7 | 43,3 | 35,9 | 43,1 | | |

In the case of the normalized data, the results were quite better, as different combinations of feature selection and prediction's results improved in comparison to raw data. There are few combinations of algorithms for which RMSE is under 42 (which was the lowest score with the raw data). Also, it needs to be highlighted that an aggressive algorithm such as CFS that chose very few variables had the ability to retain most of the information, with the almost the lowest mean value and the overall lowest score. This score was reached with the Bayesian Regression, which seems to perform better in datasets with fewer variables.

Regarding the top features selected by the algorithms (in both normalized and initial datasets), it can be remarked that vibration related ones were the most important, mostly those ones belonging to frequency domain. Skewness and median were elected more than once, meaning they are relevant for prediction. In relation to the harmonics, first and second one seemed to have similar importance and greater than third, fourth and fifth. Also, the maximum value of the harmonics was determinant and the 30 Hz RMS window overcame the 10 Hz RMS one, which could be further analysed in future works.

Respect to the rest of the variables and according to the results, sound, intensity and encoder signals have lower importance but yet not negligible. They were not repeated many times amongst the top variables. However, it was possible to find a couple of standard deviation, median and shape factor among those top descriptors.

The rest of the predictors did not seem to have as much prediction value as the previously mentioned. Not even pressure angle, which could imply, that slight changes in the morphology of the gears are not related to pitting phenomenon.

4. DISCUSSION AND FUTURE WORK

As found in the literature, feature selection algorithms emphasize the influence of the descriptors extracted from vibration signal to detect and predict pitting. They show the best results when using those obtained from harmonics in the spectrum. However, microphone, current and rotation speed related descriptors also appear to have some importance although they had less relevance for prediction algorithms. Anyway, they cannot be discarded lightly. Specific descriptors such as harmonics of vibrations have shown good performance in the case of vibrations, but the energy index has not shown promising results, possibly due to the quality of the microphone and lack of good filters. Interesting opportunities for the future would be adding more specific descriptors in the frequency or time/frequency domain for current and microphone signals to see whether they might improve the prediction of the pitting and make the sensor comparison fairer. Because right now almost all the variables were copied by vibrations, due to the knowledge in this area which supports very specific descriptors in frequency domain (side-bands).

In relation to the algorithms selected to test the subsets, it is clear that those belonging to decision tree family are better adapted to this particular case of regression. Even if the best results were reached with the

Bayesian Regression, as it needs datasets with very few variables to work properly. Nevertheless, the results are still poor and must be improved in the future. In this sense, other prediction techniques and algorithms such as Neural Networks could be tested as well as advanced Time Series analysis techniques. Those techniques have the advantage of providing the chance to execute double class regression (for wheel and pinion) which has been one of the mayor obstacles of this work.

Last, the necessity of normalizing the data must be mentioned, as it has been able to improve almost all errors, considerably reducing the error even in raw data. This technique of normalization has helped to decrease the error and therefore increase the prediction accuracy by reducing the noise of signals and slight differences among tests.

All in all, there is work to be done in the field of early pitting detection in gears, and it seems vibrations signals will play a key role on it, but without leaving aside other signals, which have also great potential.

REFERENCES

- [1] C. Hu, W. A. Smith, R. B. Randall, and Z. Peng, "Development of a gear vibration indicator and its application in gear wear monitoring," *Mech. Syst. Signal Process.*, vol. 76–77, pp. 319–336, 2016.
- [2] D. Mba and R. Rao, "Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating," *Shock Vib. Dig.*, vol. 38, no. 1, pp. 3–16, 2006.
- [3] K. R. Al-Balushi, A. Addali, B. Charnley, and D. Mba, "Energy index technique for detection of acoustic emissions associated with incipient bearing failures," *Appl. Acoust.*, vol. 71, no. 9, pp. 812–821, 2010.
- [4] I. Bravo-imaz, H. Davari, Z. Liu, A. García-arribas, A. Arnaiz, and J. Lee, "Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging," *Mech. Syst. Signal Process.*, vol. 94, pp. 73–84, 2017.
- [5] E. Shipley, "Gear failures," *Mach Design*. p. 12, 1967.
- [6] Kkaarthic, "Failure Modes in Gear: Part-1," *Bright Hub Engineering*. [Online]. Available: <http://www.brighthubengineering.com/cad-autocad-reviews-tips/8443-failure-modes-in-gear-part-one/>. [Accessed: 01-Jan-2017].
- [7] C. M. Lewandowski, N. Co-investigator, and C. M. Lewandowski, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning," *Eff. Br. mindfulness Interv. acute pain Exp. An Exam. Individ. Differ.*, vol. 1, pp. 1689–1699, 2015.
- [8] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.
- [9] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," *Mach. Learn. Proc. Fourteenth Int. Conf.*, vol. 5, pp. 296–304, 1997.

Active Lubricant Condition Monitoring

Jesvin George, Dr Ahmed Onsy
School of Engineering, University of Central Lancashire
Preston, Lancashire, PR12HE, United Kingdom

ABSTRACT

Lubrication in a mechanical system is like blood flowing through a human body. It is a great source to detect in advance any troubles in the system. As blood analysis shows the diseases in our body in a machine the oil analysis shows the fault in a system. Active lubricant condition monitoring is a great prediction technique which would help improve the reliability and reduce the maintenance costs for a system. The paper discusses about how to develop an Active oil condition monitoring where oil condition can be monitored online and how to achieve an active maintenance action. It discusses the design and setup of an oil condition monitoring system on a test bench provided. Different oil degradation parameters are analysed and sensors are introduced to monitor this parameters. Integration of microcontroller for wireless communication and use of different software's to acquire and process the data from the sensors to provide the real time condition of the system is discussed. This includes different actuation system that can be introduced to help the maintenance of the system to minimum and reduce the maintenance or damaging of the system components.

Keywords: Active Condition Monitoring System, Lubricant Monitoring, Oil Degradation, Online Oil Debris Analysis, Maintenance engineering technologies and integrated systems

Corresponding author: Jesvin George (Email: jgeorge@uclan.ac.uk)

1. INTRODUCTION

Condition monitoring is the process of monitoring or determining the condition of different parameters of a machinery while it is running. It is a part of preventive maintenance technique which enables to do maintenance of the faulty elements and prevent any unwanted failures. This kind of maintenance technique ensures the operation of the machine is going smoothly and efficiently for a long period of time thus creating a more reliable system.

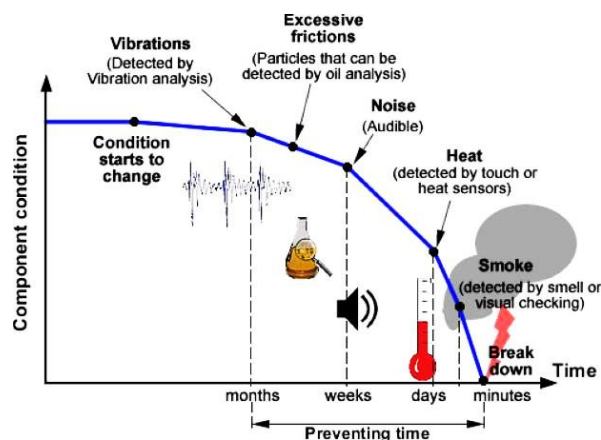


Figure 1 Development of a mechanical failure [20].

There are different kinds of condition monitoring techniques applied in industrial sectors. Some of the main techniques are: Vibration analysis, Lubricant Analysis, and Acoustic emission. Vibration analysis is a key

component in condition monitoring. It is most commonly used in detecting faults in moving elements in a machinery such as motors, bearings and gears etc. [1]. It can detect any misalignment, unbalance and resonance due to a damage in a machine. Vibration frequencies from each component are displayed in Fast Fourier Transform spectrum. Each faults generate unique vibration frequency patterns which can be used to identify the faults. Different types of sensors are used for analysis depending on the vibration frequency e.g. accelerometers for high frequency, position transducers for low frequency [2].

Acoustic emission technique is a new technique in the field of condition monitoring. It uses high frequencies of the transient elastic waves generated by rapid release of energy from a localised point from a material. Acoustic emission occurs at a frequency range of 1kHz to 2MHz or greater [3]. The elastic waves are picked up using AE sensors, converted into electric signals and amplified using a pre amplifier to produce a time waveform spectrum. Ultrasound transducers can be used to pick up AE signals.

Timely lubrication is one of the essential process required for the smooth running of any mechanical system. Any moving parts in a mechanical system is lubricated to reduce friction between the moving elements. This will reduce vibration, acoustic disturbances, wear and tear of the system. The system needs to be actively lubricated to achieve this. The maintenance frequency of the system can be significantly reduced by introducing a technique to monitor the lubrication fluid regularly and from the analysis of this fluid the fault with the system can be deduced [4]. The temperature, quantity level, vibration and the debris within the lubrication fluid can be monitored and examined regularly using different type of sensors and the problems can be determined by the information provided by the sensor [5]. This type of condition monitoring system is much useful in equipment or machineries that are not accessible easily for maintenance or of high maintenance cost. E.g. offshore wind turbines, hydroelectric generators.

Active Lubricant condition monitoring is a significant means to ensure a proactive maintenance for an equipment [6]. It is a method to ensure the lubricant is in its acceptable level and condition to meet its function but to obtain information on the condition of the equipment. It also deduce any unwanted failures, anticipate any problems thus avoiding an expensive maintenance and extend the lifetime of the equipment [6] [7]. Lubricant also has its own life time; various factors contributes to lubricant fluid degradation. High temperature, oxidation, water content, foaming and debris accumulation are some of the reason for lubricant degradation. Monitoring some of the common parameters can give the sufficient data about the lubricant condition and about the equipment. Oil degradation is the term used to describe deterioration of physical and chemical properties of the lubrication oil while it is in service. It can affect the fluid performance and diminish the service life of the oil. This can also lead to the failure of major mechanical systems, making it more susceptible to maintenance works and less reliable. Understanding the mechanisms of oil degradation and methods to rectify it can help to increase the service life of the machine. The main factors affecting oil degradation are Oxidation, water contamination, aeration, thermal degradation and wear debris particles [8]. Understanding this factors and their root cause can be used to develop a sensor-based system that can monitor the running condition of the machinery and maintain it proactively. Different sensors can gain information on this factors and the information can be processed using advanced software to relate with observed results to provide the condition of the lubricant and equipment [4]. This can provide a time frame for the required maintenance and measures can be taken to prevent any maintenance.

Oxidation is one of the main cause of oil degradation. A chemical reaction can alter the composition of the oil over time. It is result of exposure of the oil with oxygen under high temperature. Even though additives are added to prevent oxidation, it can be depleted due to high temperature. Oxidation of the oil is the main cause of problems such as viscosity increase, rust formation and corrosion of metal components in machines [9].

Lubrication oil has small percent of air present in it. It can be in form of dissolved air, free air or as foam. These forms of air presence in the can create air bubbles in the circulating oil. The dissolved air under high temperature can cause oxidation of the oil. Air bubbles can create cavitation within the metal components

[10]. It can also cause component wear due to reduced lubricant viscosity [11].

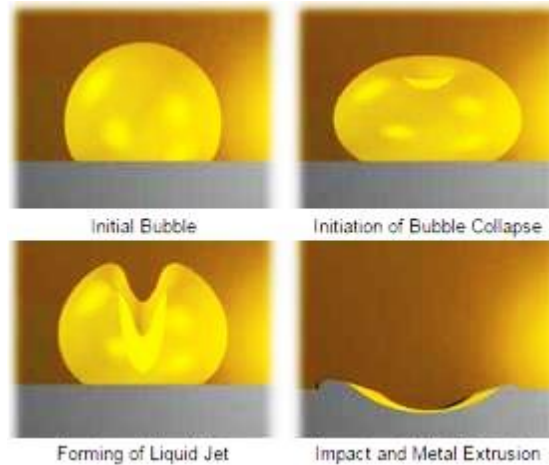


Figure 2 Cavitation in metal due to air bubbles [11].

Water presence in oil can also cause oil degradation. Water can find its path into machine parts and be dissolved or suspended in oil. This can lead to increased depletion of additives in oil and hence causing oxidation. Water content in oil can be exposed to parts in motion and cause hydrogen embroilment that in turn reduces fatigue life of metal parts. Excessive wear can occur in rotating parts because of loss of the hydrodynamic oil film due to the incompressibility of water compared to oil. For e.g. life of a journal bearing can be reduced as much as 90% as result of water content as little as 1%.

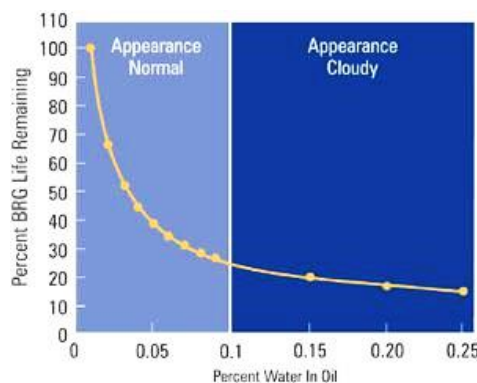


Figure 3 Graph showing relation between life of a bearing and water percent in oil [18]

Thermal degradation is the term used to describe oil degradation due to high temperature. Temperature can affect the characteristics of the oil significantly. High temperature can cause oxidation and change the viscosity of the oil. Low temperature can affect the pour point of the oil. These can change the flow of the oil resulting in insufficient lubrication of the components and damage to the components.

Wear or Debris particles of microscale can get accumulated in the oil and change the oil characteristics. Debris accumulation can occur when wear particles are washed down by the oil from machine parts. This debris when come into contact between any parts in motion can cause further wear and thus creating a path for failure of the part. Wear can be caused due to different reasons. Material removal on a surface of an object depends upon the load acting on it, composition of the surface and nature of the stress-inducing load and the environmental conditions. Understanding different types of wear and nature of it can help to identify the problem within a working machine. Generally, wear modes can be classified into three types

1. Surface to Surface wear
2. Fluid to surface wear

3. Environmental to surface wear

Understanding different types of wear and nature of it can help to identify the problem within a working machine [19].

Rubbing (Break in) (Abrasive Wear), figure 4 (d), is the most common type of wear. This type of wear occurs in an equipment when a sliding element rubs with a stator surface or in between two sliding elements. This results in the formation of low wearing surface with particles of benign nature in the form of platelets. Wear debris are mostly found in the lubricants during machining

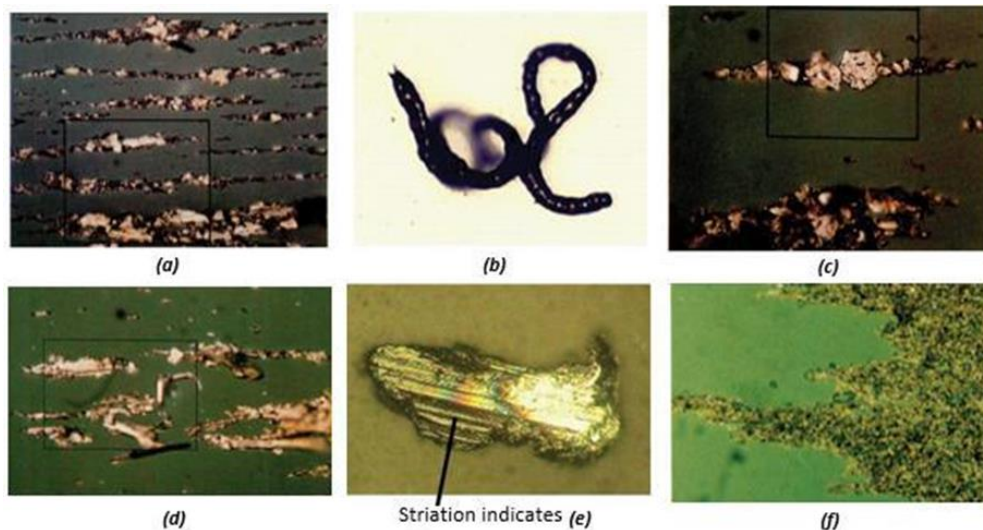


Figure 4 a) Rolling and sliding wear debris (b) Cutting wear debris (c) Rolling wear debris (d) Rubbing wear debris (e) Sliding wear debris (f) Chemical/corrosive wear debris

Cutting Wear (Abrasive Wear), figure 4(b), is produced when two surface penetrates with each other. The debris is formed when one surface is gouging over the other resulting in long ribbon like pieces. eg: Machining swarf on a lathe. Rolling Wear (Surface Fatigue), figure 4(c), produces fatigue on the surface and found in the components of rolling motion contact. Debris formed in the wear are in spherical or laminar shape. eg: Rolling motion contact in ball bearings. Sliding wear, figure 4(e), occurs due to the impact of load and speed between the contacting surfaces. The contacting surfaces breakaway small to large particles. As the load and speed between the contacting surfaces increases, size of the debris and wear rate increases. Rolling and Sliding combined wear, figure 4(a) is caused by the combined action of fatigue and scuffing. This wear is commonly found in gear system, particularly in the gear line (between two gears). The combination of sliding and rolling produce a complex fusion of the debris in between the gears (contact surfaces). Chemical/Corrosive wear, figure 4(f) is caused by fluid properties or heavy contamination from water, acid, bacteria or acid, which results in the formation small particles. In addition to this impurity from the environment also influence in this type wear. To tackle this type of wear lubricants are added with additives.eg: ferrous oxide particles formed due to the reaction between oxygen and iron.

2. DEVELOPMENT OF OIL CONDITION MONITORING TEST BENCH

The test bench for the project was assemble and provided by the University. The test bench is by the size of 1m to 0.7m size. It includes of an oil circulation system consisting of oil tank, PVC pipes, different sensors, centrifugal pump and filter. The test bench is updated with different sensors to monitor the oil condition, temperature, humidity, Viscosity, wear and Debris particles. The actuator system is also updated to respond to any undesirable condition by installing a 3-way valve that can change the flow of to the machine to avoid any further damage.

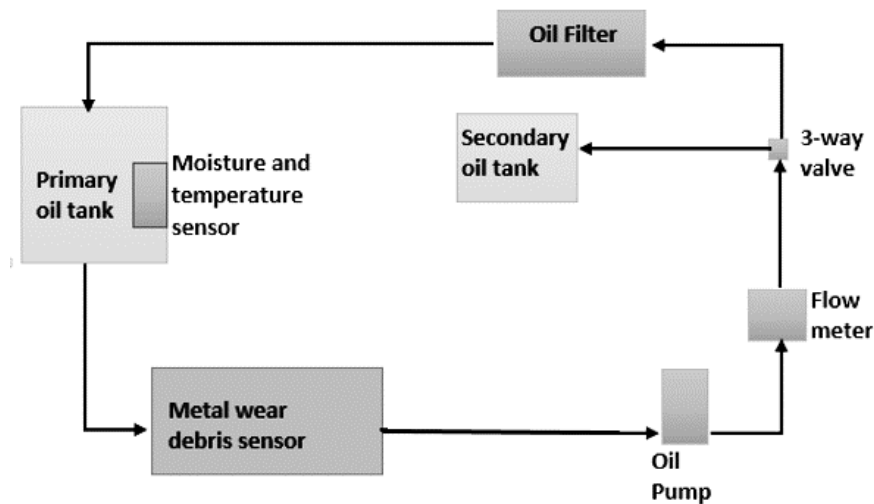


Figure 5. Block Diagram for test bench design.

The oil is stored in the tank and flows through the PVC pipe for lubrication as required. The PVC pipe is connected to a Metallic wear debris sensor and a flow rate sensor. The Debris sensor assess the contaminant particles in the oil and the flow rate sensor checks the flow rate of the oil passing through the sensor. The oil is then streamed into a 3-way valve which determines flow of the oil according to the debris data from the Debris sensor. The 3-way valve is used to change the direction of oil flow to a different tank if oil contamination or degradation is higher than the acceptable value. If the contamination or degradation of oil is on acceptable levels, the oil direction is not changed and it is filtered using a filter system. The filtered oil then flows back to the tank where the temperature and moisture sensors are placed and assess the oil condition. The sensors are interfaced with the LabVIEW software using a wireless microcontroller. The microcontroller system is used for data acquisition and actuation signal output. A wireless communication is setup over Wi-Fi with the PC.

A 10 litre stainless steel fryer is used as the oil tank in the test bench. It also includes a heater that can be used to heat up the oil to a temperature of 120°C. Two different openings are placed on the oil tank where the PVC pipes for the oil to flow out and Flow in are connected.

PVC pipes was used for piping of the system. PVC pipes are lightweight, flexible and has a main advantage on application of joint tightness. It is flame resistant to temperature 450°C. Its elastic properties reduces the magnitude of pressure surges. The pipe has a diameter of 15mm and has different fittings including angle, male and female fittings with O – ring oil seal to avoid any oil leakages due to pressure.

Filters used are a stainless steel cleanable with magnetic filter element to collect debris of microscopic range. In order to increase filter corrosion resistance, the filter surface has been anodized with bright dip blue finish. The filter is supported by an O- ring.

3-way valve is a 3 Port BSP ½ 2 DN15 Brass valve by BACOENG. It is an electrical valve that can change the direction of the flow of oil from main valve to secondary valve. A PWM signal from a microcontroller can be used to achieve this.

3. DESIGN AND DESIGN AND DEVELOPMENT OF AN ACTIVE LUBRICANT CONDITION MONITORING SYSTEM

For effective lubricant condition monitoring, a low cost sensor system needs to be developed to monitor different oil degradation elements. The system is to be designed to identify the faulty component and

property of damage depending on debris size and characteristics.

Viscosity of a liquid can be explained as a quantity used to describe as fluids resistance to flow. It is an important property of a lubrication fluid. Oil has lower viscosity than water as it less dense that is oil have less resistance to flow than water. Change in viscosity of lubrication fluid can be due to oil degradation. Oil degradation can increase or decrease viscosity of the lubrication. Increase in viscosity can caused due to oxidation or due to contamination in oil. High viscosity can restrict the flow of oil to components and reduce the ability to lubricate. Decrease in viscosity of oil can cause weakening of the oil and will not be able to prevent contacts between metals. Monitoring the viscosity of the Lubricating oil can provide us information about oil degradation and condition of the equipment. Comparing the data acquired to the known parameters can help to identify the problems. Viscosity of a fluid can be calculated using the Poiseuille's Law [12].

$$Q = \frac{\pi Pr^4}{8\eta l} \quad \text{Equation 1}$$

Where;

Q is the Flow rate

P is the change in pressure in the valve r is the radius of the valve

η is the Viscosity

l is the length of the valve

In the test bench provided a flow meter is installed to acquire the flow rate of the oil and an oil pump is used to circulate the oil around the system. The pump provides a pressure of 0.8bar which is 80kPa. The radius of the piping is 7.5 mm and the length is the distance between the pump and the flow meter. Using these values, the Viscosity of the oil can be determined.

Flow meter used is a flow rate sensor that measures the flow of oil through the system. It has a rotor, Hall Effect sensor and PCB electronics. The flow of oil can be measured by connecting it to the micro controller. The microcontroller outputs the flow measurement by counting the pulse rate of signal output from the flow meter. It works up to temperature condition of 80°C.

The centrifugal pump used to circulate the oil is a Merry Tools Mini- type pipe pump. It is a 90W 0.8 bar electronic water pump booster. It has a stainless steel motor housing, brass impeller and copper wires. The pump can be activated by switching on manually. The pump can be used to for a liquid temperature of up to 80°C at a max flow rate of 18L/Min.

A metallic wear Debris sensor is used to determine the size and quantity of the wear and debris particles present in the oil. Parker kittiwake On-line metallic wear debris sensor provides real time wear debris count for both ferrous and nonferrous wear metals. It uses magnetometry combined with smart algorithms to provide a particle size speculation and count. It can detect ferrous materials of size greater than 40 micron and non-ferrous particle of greater than 135 micron. It can quantify the metallic composition size category and particle count of the debris in oil. It operates on a temperature of -20°C to 70°C [13]. The sensor has a fluid flow range of 0.3 – 1.9ms⁻¹. The minimum flow rate and maximum flow rate required for the sensor can be thus calculated.

$$\text{Flow rate } Q = V.A$$

Where V is the volume of fluid and A is the cross sectional area of the pipe $A = \pi r^2$

$$\text{Min Flow rate } Q = V.A$$

$$= 0.3 \pi 0.0075^2$$

$$Q = 0.000053\text{m}^3/\text{s} = 3.18\text{litres}/\text{min}$$

$$\text{Max flow rate } Q = V.A$$

$$= 1.9 \pi 0.0075^2$$

$$Q = 0.00035\text{m}^3/\text{s} = 21\text{litre}/\text{min}$$

The pump provides a maximum flow rate of 18L/Min which falls within the range of sensor. The sensor is capable of detecting Ferrous and non-ferrous particles. It can also measure the size of the particles. It works on a balanced coil, full loop system. Three equally spaced coils are wrapped around an aperture. The transmitter coil in the centre creates an electromagnetic field. When the debris pass through aperture there is a disturbance in the magnetic field strength and the disturbance is picked up by the receiver coils and analysed by controlled electronics. Detection occurs if the sensitivity threshold is exceeded.

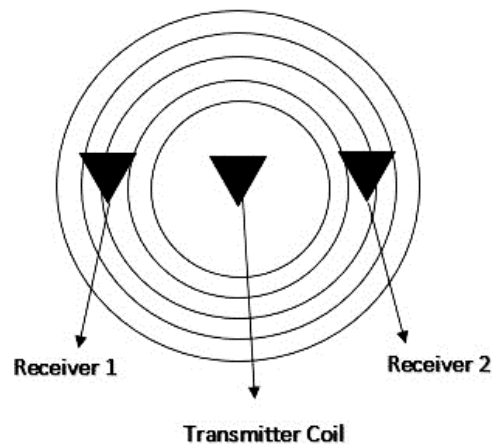


Figure 6 Schematic diagram for aperture of Wear debris sensor

The sensor assumes the debris spherical and calculates its volume using its dimensions. The spherical debris diameter is used to calculate the comparative volume of wear debris and pinpoint the origin of debris. A NTC temperature sensor was used to detect the temperature variations in the tank. The sensor has a temperature range of -40°C to 120°C . The thermocouple is used as resistor in a voltage divider circuit. So as the thermocouple value changes the output voltage changes. 10k ohm resistor is used as the thermocouple has a resistance of 20k ohm. To calculate the amount of water content in the oil a soil moisture sensor controller module was used. It senses the humidity in the environment. It has a potentiometer to adjust the sensitivity and is used to produce a warning if the water content in oil rises certain level.

NI LabVIEW is a development environment with a graphical programming syntax to create and code engineering systems. LabVIEW is integrated with a wireless Atmel microcontroller system to acquire the measurements from the sensor. Measurements are analysed in LabVIEW and control system can be developed for the automation of the system. Modbus is a communication protocol developed to be used with Programmable logic controllers. It is a communication protocol designed to transmit data between electronics devices. Modbus has both TCP/IP and serial Communication to transfer data [21]. Modbus TCP/IP uses TCP interface running on Ethernet. The MWDS uses Modbus communication to transfer the data to PC. The Modbus Register Map for the MWDS can be accessed by developing a program in LabVIEW.

The system can be programmed to send a notification to the Engineer in case of a failure. It can be designed to send a notification email from LabVIEW using the SMTP email function. The user email information can be configured by entering the senders email address and outgoing mail server. DebrisSCAN is a software

that can be used to acquire the data from MWDS. The software provides the different characteristics data of the debris in oil and gives feedback and alert whether the system is under normal running condition or abnormal running condition.

The Atmel microcontroller (IDE) is a software used to write the Atmel microcontroller programs and upload them into the Atmel Board. The Atmel IDE has example program for each board which is helpful to write the program for the system. It can be programmed to read the sensor data and display it in serial monitor. Using Atmel microcontroller the Uploading the program and displaying the data in serial monitor can be done wirelessly using Atmel IDE.

Open SSH is an application available in the android operating system that allows to connect to the Atmel microcontroller using Secure Shell network protocol. The SSH protocol allows to connect to a server securely through an unsecured network. The Open SSH reads the data send by the Atmel and displays it in the remote device like a serial monitor.

An intelligent system needs to be developed for the system adapt to the different environmental or machine condition e.g. heating or cooling of oil, provide or more less lubrication and communication with another unit and compare the system condition. Intelligent system can be defined as an embedded, internet connected computer machine that is capable of acquiring, analysing data and communication with other systems. An intelligent system should include security, connectivity, remote monitoring and management capacity and be able adapt according to the acquired data [16]. For the system to respond to a data a logical system needs to be integrated into its programming. Fuzzy logic is a logical system that can be used for fusion of different sensory measurements and create a decision making algorithm. Fuzzy logic uses the degrees of truth instead of the usual Boolean logic of true or false. It helps to give a value to the true or false using various states in between. Fuzzy logic is helpful in development of an artificial intelligence in the software that can conclude to a human like decision based on collective data from the system.

4. ACTIVE CONDITION MONITORING SYSTEM VALIDATION

The different sensors placed in the test rig was connected to the Atmel microcontroller. The code was developed in LabVIEW to acquire the signal and visual warning aids were designed to ensure the system is on normal operation limits. Preliminary testing for the sensors was done and the results was observed. The data from the temperature sensor displayed in LabVIEW using graphics. The sensor was tested at room temperature.

The LEDs was used as visual warning system for the user. It High temperature lights up if the temperature exceeds 80 degrees and low temperature led if temperature falls below 10 degrees Celsius. The moisture sensor data was acquired and displayed in LabVIEW. The sensor was tested in different conditions. It was immersed in oil, mixture of oil and water, and just water. In oil the value was above 1000 which shows dry condition and in water the value was less than 820. Unfortunately, the mixture of oil and water reading was not reliable as water droplets was covered with thin film of oil which produced a dry condition result. Significant amount of water has to be added for the sensor to show a moisture presence in oil. The oil pump was turned on and acquired the flowmeter data. The code was developed to count the pulse output from sensor. The Pulse frequency and the flowrate was calculated from the pulse count and displayed. Low flow rate limit set for 3.2litres/min and high flowrate was set for 21 litres/min. The wear debris sensor data was displayed in LabVIEW software when wear and debris particles was introduced to the oil. The MWDS was able to detect the particles and categories it into size and ferrous or non-ferrous class.

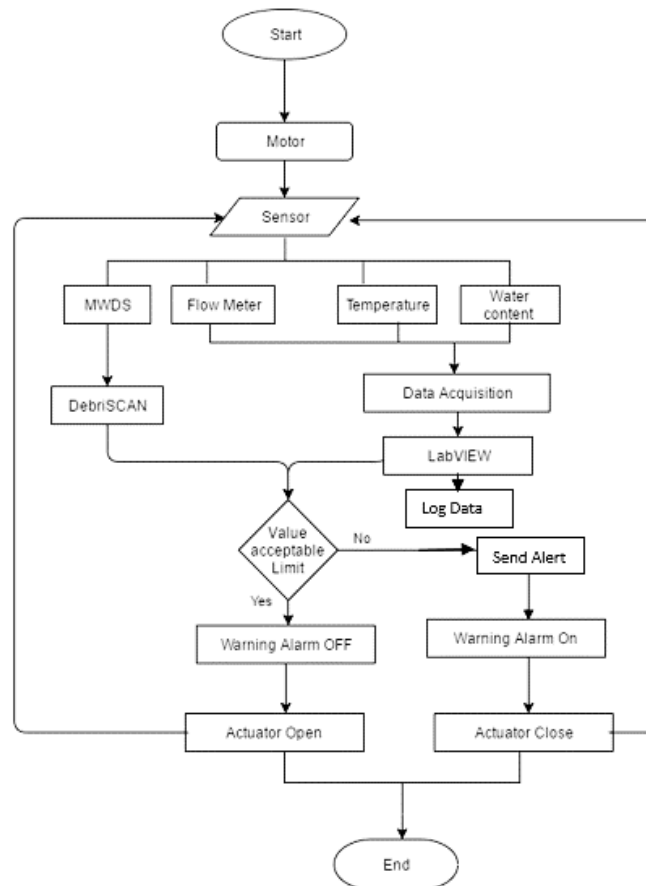


Figure 7 Flow Chart Diagram for ALCM system

The complete Active Lubricant Condition Monitoring system is a sensory actuation system controlled by an embedded system and software developed in LabVIEW. The system is developed as a machine health monitoring system which can predict the failure and the current condition of the system based on the oil degradation parameters. The system monitors the sensor inputs and activate the actuation system based on this data. The wear debris sensor detects the wear particles flowing through the system and the data is displayed in the LabVIEW interface in the relevant section. A warning limit and alarm limit can be set for maximum particles per minute and maximum mass for particles per hour can be set. When the limit is reached the system activates the actuation system and changes the direction of the flow of oil from primary tank to the secondary tank using the 3 way valve. Once the particle count per minute or mass per hour values goes down below the set limit, the system activates the actuation system changing the flow of oil into primary tank to default operation. The control of the actuation system is automated based on the system condition. It can also be manually controlled using a flip button in case any malfunction to occur and control of the actuation system is required.

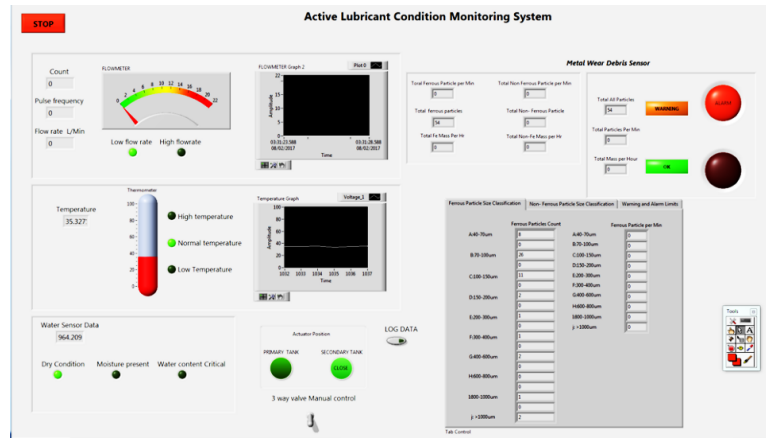


Figure 8. Active Lubricant Condition Monitoring system in LabVIEW

The wireless Active condition monitoring system was developed using Atmel IDE. Sensors was connected to Atmel microcontroller board. The Yun microcontroller board can be connected to the PC using Wi-Fi connectivity. This allows the microcontroller to send and receive the sensory actuation system data wirelessly.

The 3 way electric valve was able to be controlled wirelessly through Atmel microcontroller and Notification for is displayed on the serial monitor when the valve opens or closes.

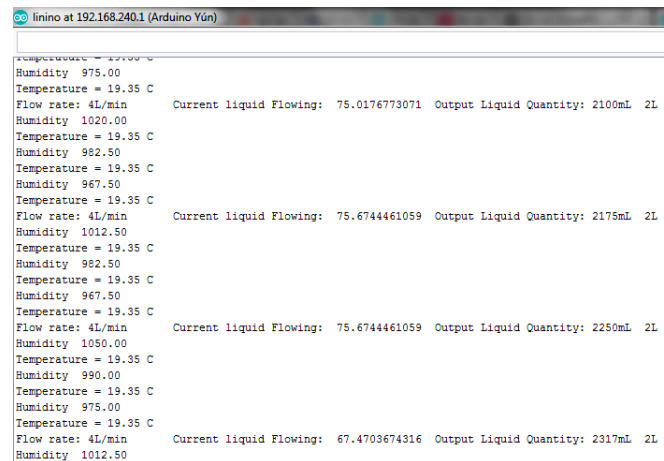


Figure 9 ALCM system data displayed in PC wirelessly

The Wireless ALCM system can be accessed wirelessly through any PC, Smart phone or tablet as long as it has Wi-Fi connection to connect to Atmel microcontroller and Serial program or SSH application available in the device. The actuator can also be controlled through this device as same as the PC to close or open the 3 way valve. Security of the system was considered. The system is also secured using password. The system can only be accessed through SSH client application in smart devices. SSH is a Secure Protocol used primarily as means to connect remotely to Linux servers.



Figure 10 Mobile SSH client application with the shared data displayed on a smartphone (Android platform)

5. CONCLUSION

The ALCM Program was developed in LabVIEW software to monitor and control system in real time. The aim of the project was accomplished by monitoring the factors that causes oil degradation. MWDS, Temperature sensor, Water Content sensor, Flowmeter are the sensors used to monitor the oil conditions. The sensors were able to provide information regarding the oil degradation parameters such as particle count, classification, size of debris, water content, temperature and viscosity. The actuator system was designed using a 3- way valve, it was able to direct the flow of oil to a secondary tank when debris particle of set limit is detected in the system. This design feature was included to avoid any further damage to the system due to the degraded oil. The system was able to alert the user of system failure by sending an email. An ALCM program was also developed using Atmel microcontroller and software interface to monitor the system condition and control it wirelessly. The sensors were tested and a measurement was acquired but more research and testing has to be done to interpret the data and recognise significant characteristics machine condition so that; it can also provide a most likely scenario of failure where potential consequences and its probability of occurrence can be determined. Active real time monitoring of the system can provide any indication to potential failure and condition of the system.

6. FURTHER WORK

Fusion of measurement from these sensors using a fuzzy logic system can provide an information of condition of the machine to create an intelligent ALCM system. The system can be made more efficient elemental analysis is required to identify the failed components. Further testing of the system with different failure modes is required and develop the system to identify the root cause of the failure. Combine oil condition monitoring with other condition monitoring techniques such as vibration and acoustic emission monitoring to design an effective condition monitoring system. The Atmel controller has the functionality to connect to the internet which can be used to send and receive data through TCP/IP. This can help to create an ALCM system that can be accessed via online. Interfacing the Atmel microcontroller with LabVIEW to create a GUI which is easier for the user to interpret the Sensor data. Develop an intelligent Maintenance system with sensors and actuators were the Machine can predict the failure and use the actuators to rectify the failure or inhibit any further damage to the system. Develop an ALCM system that can communicate to other system, compare the data to self-assess the system performance and make it more efficient and reliable.

REFERENCES

- [1] Peters JD. Beginning vibration analysis with basic fundamentals. [place unknown]: Vibration Analysis hardware; 2005.
- [2] Tiwari R. ANALYSIS AND IDENTIFICATION IN ROTOR-BEARING SYSTEMS all India Council of technical education. Indian Institute of Technology Guwahati: All India Council of Technical Education; 2010.
- [3] Mark F. Carlos. standardization news. [place unknown: publisher unknown]. ASTM acoustic emission standards [cited 2016 Sep 14]. Available from: http://www.astm.org/SNEWS/OCTOBER_2003/carlos_oct03.html.
- [4] ONSY, A., BICKER, R., SHAW, B. A., Intelligent Diagnostic Health Management of Power Transmission Systems: An Experimental Validation. 2010 July; 13(3).
- [5] Dempsey PJ. Integrating Oil Debris and Vibration Measurements for Intelligent Machine Health Monitoring: National Aeronautics and Space Administration; 2003.
- [6] Abel Resina de Almeida GE. Lubricant condition Monitoring. 2003 June 5: p. 231- 243.
- [7] Paresh Girdhar CS. Oil and particle analysis. In Scheffer C, editor. Practical Machinery Vibration Analysis and Predictive Maintenance.: Newnes ; 2004. p. 168-220.
- [8] C.C.Jensen, Oil degradation, http://www.cjc.dk/fileadmin/user_upload/pdf/CJC_Brochures/Oil_Degradation_Varnish.pdf
- [9] Corporation N. machinery lubrication. Oxidation - the lubricant's nemesis; 2007 May [cited 2016 Sep 14]. Available from: <http://www.machinerylubrication.com/Read/1028/oxidation-lubricant>.
- [10] Duncanson M. Machinery lubrication. [place unknown]: A noria publisher. Controlling oil aeration and foam; 2001 [cited 2016 Sep 14]. Available from: <http://www.machinerylubrication.com/Read/255/oil-foam>.
- [11] Hi NC, Tun VQ, Th T, Hà T, Ngc N, Minh, Trinh CC. Fluidic Capacitive sensor for detection of air bubble inside engine Lubricating oil. VNU Journal of Science: Natural Sciences and Technology. 2015 [cited 2016 Sep 14];31(1):8–16. Available from: <http://tapchi.vnu.edu.vn/upload/2015/05/1687/2.pdf>.
- [12] FLUID FLOW VISCOSITY POISEUILLE'S LAW. Sydney: VISUAL PHYSICS School of Physics University of Sydney; 2014.
- [13] Parker Kittiwake. [place unknown]: Parker Kittiwake. On-line metallic wear debris sensor; 2015 [cited 2016 Sep 14]. Available from: <http://www.kittiwake.com/metallic-wear-debris-sensor>.
- [14] Minasi M. Ultrasonic sensing for fluid identification and contamination. Application Report SNAA265. 2015 [cited 2016 Sep 14]. Available from: <http://www.ti.com/lit/an/snaa265/snaa265.pdf>.
- [15] Minasi M. How to select and mount transducers in ultrasonic sensing for level sensing and fluid ID. Application Report SNAA266A. 2015 [cited 2016 Sep 14]. Available from: <http://www.ti.com/lit/an/snaa266a/snaa266a.pdf>.
- [16] Contributor, Wigmore I.: WhatIs.com. What is intelligent system? - definition from WhatIs.com; 2013 Oct [cited 2016 Sep 14]. Available from: <http://whatis.techtarget.com/definition/intelligent-system>.
- [17] Rouse M.: WhatIs.com. What is fuzzy logic? - definition from WhatIs.com; 2016 Aug [cited 2016 Sep 14]. Available from: <http://whatis.techtarget.com/definition/fuzzy-logic>.
- [18] Corporation, N. (2001) Water in oil contamination. Available at: <http://www.machinerylubrication.com/Read/192/water-contaminant-oil> (Accessed: 15 September 2016).
- [19] Fitch, B. (2013) Anatomy of wear debris. Available at: <http://www.machinerylubrication.com/Read/29537/wear> (Accessed: 15 September 2016)
- [20] Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T., Ekemb, G., Tameghe and Andy, T. (2014) 'Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges', *Energies*, 7(4), pp. 2595–2630. doi: 10.3390/en7042595
- [21] Modbus, S. (2015) About Modbus. Available at: <http://www.simplymodbus.ca/faq.htm>

Value-Based Management in Centrifugal Compressors Maintenance

X. Liang¹, F. Duan¹, D. Mba¹, I. Bennett²

¹ School of Engineering, London South Bank University, London, SE1 0AA, United Kingdom

² Team Lead Technology–Rotating Equipment, Shell Global Solutions International, B.V., Rijswijk 2288 GS, Netherlands

ABSTRACT

Nowadays, the condition-based maintenance (CBM) is attracting considerable interests in both academic and industrial field due to its capabilities of detecting equipment fault at an early stage, being able to schedule maintenance in advance, and avoiding costly unplanned downtime. Although CBM can effectively avoid unscheduled breakdown maintenance, it may suffer from the problem like over maintenance, which will result in unnecessary economic loss. To minimize the overall maintenance cost, a novel maintenance approach for a centrifugal compressor system is proposed in this paper by incorporating the value based management (VBM) into the CBM. In this way, an accurate CBM total maintenance cost model is built and the maintenance action will be triggered at the time when the predicted total maintenance cost reaches its minimum value. The proposed model is then demonstrated with a case study on the common fouling fault in centrifugal compressor. The results of the case study show that the cost model can save approximately nearly 0.1% to 1% of the overall maintenance cost per maintenance and reduce probably one time of off-line maintenance per year.

Keywords: condition-based maintenance, value-based management, maintenance cost model, centrifugal compressor, fouling.

Corresponding author: X. Liang (liangx3@lsbu.ac.uk)

1. INTRODUCTION

Centrifugal compressors are widely applied in a variety of industrial applications, such as oil and gas production, gas transportation and refrigeration [1]. Due to their critical roles and wide applications, unscheduled breakdown of centrifugal compressors can cause significant economic loss to a company [2]. In addition, the over-maintenance for a centrifugal compressor may also result in increased maintenance cost. Therefore, efficient maintenance strategy is essential to ensure reliable operation, avoid unscheduled downtime and thus reduce overall operational cost [3].

Over several decades development, the maintenance strategies have experienced several stages [4]. In the early days, maintenance was performed only after breakdown occurred. Apparently, this method was not suitable for some critical machines, like centrifugal compressor, whose unplanned breakdown may not only lead to huge economic loss, but also may cause environmental contaminations and personnel casualties [5]. A later and better maintenance strategy is called periodic maintenance, which sets a predefined periodic interval to perform preventive maintenance regardless of the health status of a physical asset. This maintenance approach can greatly reduce the unplanned shutdown. However, this strategy suffers from significant economic loss due to over maintenance or lack of maintenance when some equipment failure occurs within the maintenance interval [6]. Therefore, quite a few researches have been conducted to optimize the maintenance interval. For instance, Bris et.al [7] and Samrout et.al [8] optimized the inspection cost in a general series–parallel system and tried to find a more suitable maintenance interval. Nowadays, as modern machines become increasingly complicated, it is even more challenge to find proper maintenance intervals. Eventually, a more cost effective maintenance approach is desired [4].

In recent years, the condition-based maintenance (CBM) is becoming popular and has attracted considerable interests in both academic and industrial field. In this strategy, the maintenance decision is

given according to the run-time status of the machine through analysis of a series of condition monitoring parameters. Thus, it can detect upcoming equipment failure, schedule maintenance in advance, and effectively avoid unplanned downtime [9]. A recent study shows that the operating expenses can be reduced by up to 5% with the usage of CBM strategy in petrochemical industry [10].

In current research of CBM, the maintenance time point is optimized mainly based on reliability analysis [11]–[13], with few considerations of the overall maintenance cost. Many times, in order to avoid unplanned shut-down, the CBM threshold of trigger maintenance is set with relatively high reliability, which results in the waste of the residual life of parts i.e. uneconomical over maintenance. To overcome this problem, a novel maintenance approach can be established with the purpose of minimizing the overall maintenance cost.

In order to build an accurate overall maintenance cost model, the concept of value-based management (VBM) [14], [15] is introduced into the model. VBM is a management approach that has been applied to capital budgeting, management control, incentive compensation, etc [14]. It can maximize the profit of a company by considering a relatively complete list of costs, including both financial cost and non-financial cost. Compared with breakdown maintenance, periodic maintenance, and traditional CBM, the new maintenance model, which takes VBM into account, is more economical due to a more accurate maintenance cost model and directly regarding the minimum cost as one of the maintenance objectives.

The paper is organized as follows. In the next section, an approach of combining the CBM and the VBM is introduced and then, an accurate total maintenance cost model is presented in Section 3. In Section 4, the common fouling fault in centrifugal compressor is taken as an example to demonstrate the effectiveness of the proposed model and finally the conclusion is drawn in Section 5.

2. MANAGEMENT STRATEGY

2.1. Combining CBM with VBM

For companies, making appropriate maintenance plan is one of the important processes to maximize the profit [16]. There are only a few papers that discuss maintenance cost of periodic maintenance, and even fewer papers calculate maintenance cost of CBM. For example, R. Pascual et al. [17] took the labour, materials, and downtime costs into account to optimize repair, overhaul and replacement times to reduce the total costs. Aretakis et al. [18] calculated the compressor washing cost, start up and fuel costs to find the optimal number of offline washings in one year. In terms of CBM, Nina F. Thornhill et al. [19], [20] arranged the compressor washing maintenance by considering start-up, shutdown, operating costs and costs for acquiring products from external sources. Although there are a few papers considered the maintenance cost, the cost model is very simple with very limited maintenance costs. The cost of CBM can be significantly reduced if a more accurate maintenance cost model is built. .

As an efficient cost calculation and management tool, the VBM [14], [15] is popular in controlling a firm's financial operations such as capital budgeting, valuation, management control, and incentive compensation [14]. Compared with limited costs calculated in preventive maintenance, the VBM considers both financial and non-financial factors [15], which help to establish a relatedly complete and realistic maintenance cost model. The costs listed above are all embraced in financial maintenance cost in VBM concept, and non-financial maintenance cost can include material waste cost, environment cost, accident cost etc. When the minimum maintenance cost is regarded as the maintenance objective, the realistic maintenance cost model, which takes VBM into account, will be more economical.

A novel maintenance strategy of combining CBM with VBM is developed, and illustrated in figure 1. When an incipient fault is discovered, the condition monitoring system gives an indication. Then, the prediction system forecasts the reliability, fault distribution function, failure rate and remaining life, which is a similar

process in traditional CBM method. At the same time, the estimated total CBM maintenance cost is calculated by means of multiplying costs introduced in VBM concept by the predicted condition of the system. The maintenance costs introduced in VBM concept consists of both financial and non-financial costs. The financial factors include downtime cost, parts repair cost, labour cost, fuel or electricity cost etc. and non-financial factors include material waste cost, environment cost, accident cost etc. Finally, a suitable maintenance time or time range can be found with minimum maintenance cost.

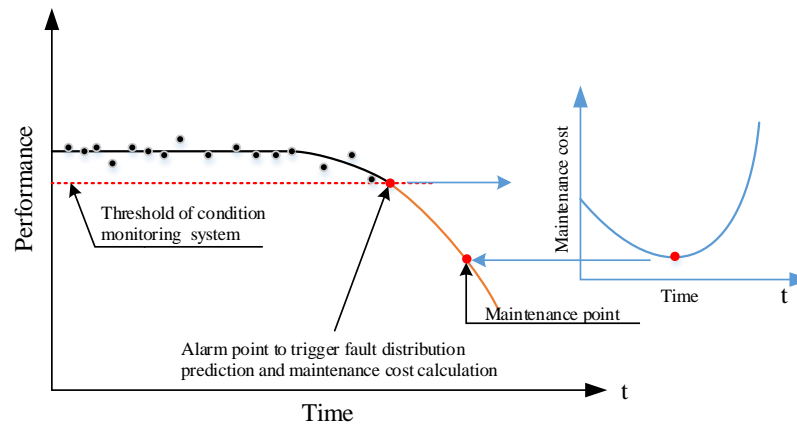


Figure 1 System degradation model and maintenance cost

2.2. Structure of centrifugal compressor maintenance system

To achieve in-time and economic maintenance, a maintenance framework is built and shown in figure 2. Initially, the failure historical data in the past several years is collected and used to calculate the time interval between failures. Then a standard fault distribution model is selected to fit the raw data. The equipment lifetime can be obtained via the failure distribution of the equipment [21]. In the process of failure distribution identification and parameter estimation, the Kolmogorov–Smirnov test, Least-Squares Curve-Fitting (LSCF), Maximum Likelihood Estimator (MLE) techniques are widely used. Using the fitted standard fault distribution model, the reliability, failure probability distribution, failure rate and remaining life can be calculated, which are used as inputs of CBM cost model. When the fault monitoring and diagnosis system detects an incipient fault, the CBM cost, combining VBM is calculated as a function of time from the fault appearance to the predicted system failure. At the same time, the estimated maintenance cost is calculated based on the predicted condition of the system.

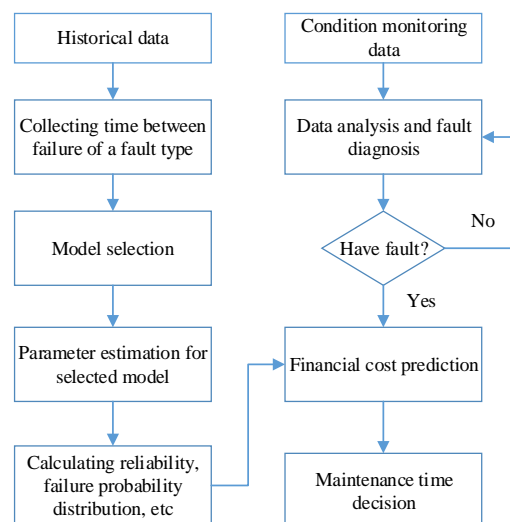


Figure 2 Structure of maintenance system

3. MAINTENANCE MODEL OF CENTRIFUGAL COMPRESSOR

The purpose of a cost model is to arrange the optimal maintenance operation by minimizing the total maintenance cost. Only considering financial costs does not provide a true picture of a company's operations [22]. From the VBM concept, total CBM costs can be expressed as the integration of financial costs, and non-financial costs, as shown in Eq. (1):

$$\text{Min } C_{CBM} = C_{Financial} + C_{Non-financial} \quad (1)$$

The validity of the model is based on the assumptions of the CBM cost model as follows:

- (1) Data of the same type of fault is collected from an identical machine.
- (2) After a maintenance action is carried out, the machine is back to normal operation without performance decrease.
- (3) Resources, tools and spare parts are always available and adequate to carry out any recommended job; otherwise, extra costs on ordering parts need to be added into the maintenance model.
- (4) Maintenance personnel is well qualified and capable to perform any sort of maintenance actions.
- (5) The power consumption due to degradation follows a linear function of the degradation rate and cumulative operational time [20]. The degradation rate of a compressor depends on the type of the compressor and on the cumulative time of operation after the last maintenance.
- (6) The effects from the surroundings do not change abruptly during a short time period.

3.1. Financial cost

The financial cost is defined as the machine cost, interest and other charges to build, purchase and install assets. In this model, financial cost in CBM includes downtime cost, maintenance cost, labour cost, fuel cost, probably down time cost if no maintenance is carried out at time t , and production benefit if compressor continues working.

$$C_{Financial} = C_{Dt} + C_{Mt} + C_L + C_E + C_{PD} - C_G \quad (2)$$

Where

$$C_{Dt} = \sum_{i=1}^k (C_{Dt\text{maintenance}}) + C_{Dt\text{startup}} + C_{Dt\text{shutdown}} \quad (3)$$

$$C_{Dt\text{startup}} = C_{dt(i,t)} \times T_{\text{startup}} \quad (4)$$

$$C_{Dt\text{shutdown}} = C_{dt(i,t)} \times T_{\text{shutdown}} \quad (5)$$

$$C_{Dt\text{maintenance}} = C_{dt(i,t)} \times T_{1i} \quad (6)$$

C_{Dt} denotes the total downtime cost, which encompasses downtime cost due to maintenance time $C_{Dt\text{maintenance}}$, downtime cost of machine start-up and shut-down time, namely $C_{Dt\text{startup}}$ and $C_{Dt\text{shutdown}}$, respectively; k is the total number of fault types; $C_{dt(i,t)}$ is the downtime costs per unit of time. The downtime of a machine results in the main profit loss of company, and can be expressed as a sum of the downtime caused by maintenance, machine start-up and shut-down. The cost $C_{dt(i,t)}$ depends on several external factors, such as production rates, stock prices, and system design parameters (redundant equipment, stock piles, and alternative production methods) [17]. The estimation value of $C_{dt(i,t)}$ can also be given by company. T_{1i} is each part's maintenance time, and it is a function of each fault type's fault distribution.

$$C_{Mt} = \sum_{i=1}^k (C_{\text{inspection}} + C_{\text{repair}} + C_{\text{replace}}) \quad (7)$$

$$C_{\text{inspection}} = N_{\text{insp}} \times C_{\text{insp}(i,t)} \times T_{2i} \quad (8)$$

$$C_{\text{repair}} = N_{\text{repair}} \times C_{\text{repair}(i,t)} \times T_{3i} \quad (9)$$

$$C_{\text{replace}} = N_{\text{replace}} \times C_{\text{replace}(i,t)} \times T_{4i} \quad (10)$$

$$T_{1i} = T_{2i} + T_{3i} + T_{4i} \quad (11)$$

The maintenance cost C_{Mt} is generated by part inspection, repair and replacement. N_{insp} , N_{repair} , N_{replace} are the number of components that need inspection, repair and replacement, respectively. $C_{\text{insp}(i,t)}$, $C_{\text{repair}(i,t)}$, $C_{\text{replace}(i,t)}$ are related to inspection, repair and replacement cost per unit of time. T_{2i} ,

T_{3i} , T_{4i} are inspection, repair and replace time, and they are all related to each fault type's fault distribution function.

$$C_L = \sum_{j=1}^L (N_{Lj} \times C_{L(j,t)} \times T_{1j}) \quad (12)$$

C_L expresses the labour costs; L represents the required maintenance staff level, for example general technician, high-level engineer and experts; N_{Lj} , $C_{L(j,t)}$ are the numbers and cost of maintenance staff, respectively; T_{1j} is related to the working hours of different staff.

$$C_E = C_E^c + \Delta C_E \quad (13)$$

C_E is the total electricity cost [20], $C_E^c = W_{E(i,t)}^c \times T_{5i}$, $W_{E(i,t)}^c$ is the power consumed when the compressor is clean, T_{5i} is the accumulation of working time.

$$W_{E(i,t)}^c = \sum (\delta_{1(i,t)} \times M_{(i,t)} + \delta_{2(i,t)} \times P_{T(i,t)}) \quad (14)$$

where, $\delta_{1(i,t)}$ and $\delta_{2(i,t)}$ are normalized coefficients of the power consumption of compressor. These parameters mainly depend on information related to cooling water consumption and ambient conditions of each compressor [20]. $M_{(i,t)}$ is the flow rate and $P_{T(i,t)}$ is the discharge pressure.

According to assumption (5) in Section 3.1, the additional power consumption ΔC_E , which is caused by system degradation, follows a linear function of the degradation rate and cumulative operational time. The power consumption due to degradation can be expressed as:

$$\Delta C_E = \Delta W_{(i,t)} \times T_{5i} \quad (15)$$

$$\Delta W_{(i,t)} = \epsilon_{(i)} \times \Delta S_{(i,t)} \quad (16)$$

$\Delta W_{(i,t)}$ is the extra power consumption. $\epsilon_{(i)}$ is efficiency degradation rate [23]. $\Delta S_{(i,t)}$ is the cumulative time of the operation after the last maintenance.

$$C_{pdown} = C_{dt(i,t)} \times T_{maintenance} \times P_{f(i,t)} \quad (17)$$

C_{pdown} is the probably down time cost, in that if no maintenance is carried out at time t , the system has a probability of suddenly shut down; $C_{dt(i,t)}$ is the down time cost per unit of time, and its calculating method has been mentioned above; $T_{maintenance}$ is the inspection and maintenance time for an unplanned shutdown; $P_{f(i,t)}$ is the failure rate function.

$$C_G = C_{g(i,t)} \times \sum T \quad (18)$$

C_G is the production benefit if compressor continues working; $C_{g(i,t)}$ is the profit per unit of time; $\sum T$ is the accumulation of time.

To get accurate cost calculation, other costs can also be added to the cost equation, such as cost for ordering and storing parts, cost for changing header [20], acquiring products from external sources and cost for installing diagnosis and monitoring system, etc.

3.2. Non-financial costs

For the non-financial costs in CBM, it can be material waste cost, environmental cost, accident cost, etc. Although the non-financial cost is usually hard to estimate, it actually occupies a large proportion of the total maintenance cost [24], [25]. Among the non-financial costs, material waste cost, which is caused by

replacing a part too early that wasting the residual life of the part, can occupy a large proportion of total maintenance costs. The material waste cost can be obtained by:

$$C_w = C_{w(i,t)} \times T_{remaining} \quad (19)$$

Where, C_w is the estimated material waste cost, $C_{w(i,t)}$ is the material waste cost per unit of time, $T_{remaining}$ is the remaining life of the system. Pollution treatment fee can be either calculated using equations in [25] or given directly by company.

4. CASE STUDY

To demonstrate the effectiveness of the proposed model, a preliminary analysis is carried out by taking the common fouling fault in centrifugal compressor as an example. The failure data are simulated and it follows the lognormal distribution [26], which is commonly used in maintenance system [27]. By Kolmogorov-Smirnov test, it shows the data follows the lognormal distribution, with shape parameter $\mu = 6.3782$ and scale parameter $\sigma = 0.53227$. Kolmogorov-Smirnov test can tell how the data is distributed, with advantage of making no assumption about the distribution of data [28]. The maximum likelihood method is employed to find the fit parameters. The result of the two parameters of the lognormal distribution is shown in figure 3(a). Then the reliability function, failure rate function, probability distribution function according to lognormal distribution are presented in figure 3(b), (c), and (d), respectively.

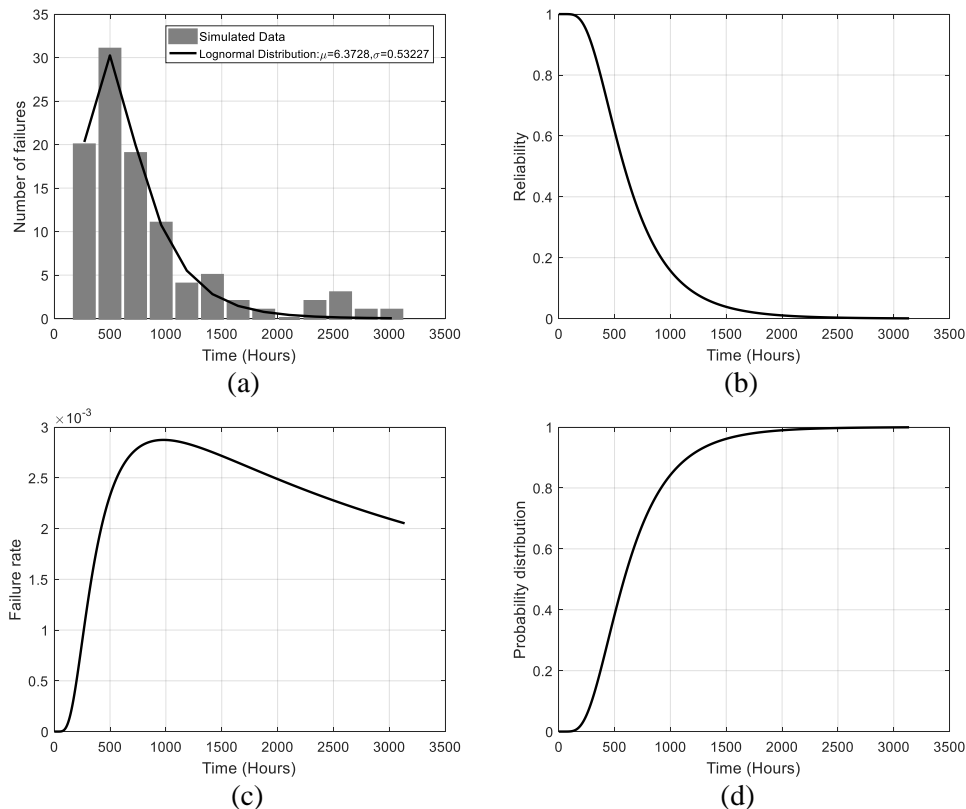


Figure 3 (a) Raw data and the fitted standard distribution, (b) reliability of the system with the given fault type, (c) failure rate function of the system with the given fault type and (d) probability distribution function of the system with the given fault type

The maintenance cost and time of the fouling fault are listed in table 1. In oil industry, on-line washing is operated once a day, and off-line washing is performed once a month. The cost of this periodic maintenance can be reduced by finding a suitable off-line washing time.

Table 1: Maintenance cost and time

| Water cost (€/h) | Electricity cost (€/kwh) | Downtime cost (€/h) | Labour cost | Estimation for material waste cost and probable shut down cost |
|----------------------|--------------------------|-------------------------------------|----------------------|--------------------------------------------------------------------------------------------------------------------------------|
| 2 €/h | 0.0984 €/kwh | the minimum loss is around 8333 €/h | About 40k € per year | minimum 3k € for buying an compressor impeller minimum 1k€ for buying big journal bearing minimum 3k€ for buying a shaft |
| Start-up time | | Shut-down time | | Maintenance time |
| At least 45 minutes | | At least 20 minutes | | incipient fouling: 2 weeks; serious fouling: 3 weeks |

Supposing fouling is detected 28 days (673 hours) after the off-line washing finished, which is an relatively early fault and its corresponding system reliability is about 0.35 (see figure 3). The CBM threshold of trigger maintenance is the reliability of 0.2, which means that the CBM suggests maintenance to carry out after 9 days (216 hours). Supposing the efficiency is fully recovered after off-line washing. By applying VBM, the estimated financial cost, non-financial cost and total maintenance cost for the new maintenance strategy is shown in figure 4.

For general compressor fouling, no high-level technicians or experts are needed, so the labour cost is constant (see figure 4(a)). The predicted downtime cost, repair cost and probability of unplanned shutdown cost all show increasing trend, as can be observed in figure 4(b), (c) and (d), respectively. Due to maintenance delay, the system degradation becomes more severe and more repair time is required. This strategy is expected to reduce costly unplanned shutdown because it takes the probability of unplanned shutdown cost into consideration. The production gain (figure 4(e)) is the profit earned if machine continues working. Its value will increase with the accumulation of working time. The material waste cost is taken in account as a non-financial cost. According to material waste (figure 4(f)), the cost decreases with the reducing component's remaining life. The total power consumption cost (figure 4(g)) increases almost linearly with the accumulation of working time. It consists of power consumed when the compressor is clean and power wasted due to degradation. With the degradation getting serious, the increasing trend of clean power consumption accumulation is slowing down, while the waste power increases, and can add up to a waste of 4%.

From figure 4, it can be observed that the highest cost is downtime cost. Based on the assumptions and parameters given by that industry, it is suggested to carry our off-line washing at 288 hours (12 days) after fouling being detected. The result in figure 4(h) illustrates that it is not cost effective to perform maintenance either at an early developing stage or a late severe stage. Compared with the periodic maintenance applied in the industry now, which operates the off-line washing once a month, the new maintenance strategy advises the maintenance action to carry out 10 days later than periodic maintenance, with a nearly 1.3% cost saving. Compared with the traditional CBM maintenance, which threshold of trigger maintenance is when reliability reaches 0.2, the new maintenance strategy advises the maintenance action to carry out four days later than pure CBM, with a 0.1 % cost saving. That implies that the maintenance strategy proposed in this paper is able to avoid the over-maintenance to some extent. According to a report by Siemens company [29], a centrifugal compressor in gas turbine may experience off-line maintenance 4-6 times per year, if this maintenance strategy is applied, the off-line maintenance can be reduced about once per year. According to the company's data, its minimum loss is around 8333 €/h. If the off-line maintenance can be reduced about once per year, it will save nearly millions of euros per year.

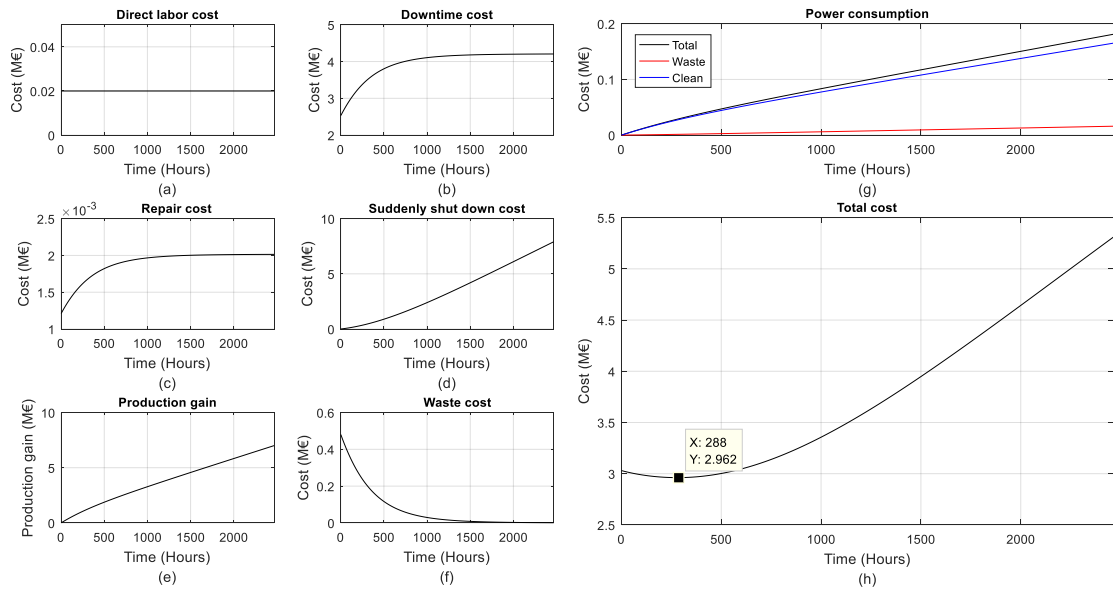


Figure 4 CBM cost when considering fouling in centrifugal compressor

5. CONCLUSION

An improved CBM maintenance strategy is proposed in this paper by introducing VBM concept into CBM. On this basis, an accurate total maintenance cost model is built, which allows the maintenance action to be conducted at the time period with minimum total maintenance cost. This strategy is expected to reduce costly unplanned shutdown and thus avoid over maintenance. The total maintenance cost includes a relatively complete list of financial costs, such as downtime cost, maintenance cost, the labour cost, electricity cost, probably down time cost and production gain. Moreover, the non-financial cost, like material waste, is also taken into account. To prove the effectiveness of the improved CBM maintenance strategy in reducing cost and avoiding over-maintenance, a case study is carried out by taking compressor fouling as an example. The results indicate that the new model could reduce the number of maintenance in a year and save nearly millions of euro per year.

REFERENCE

- [1] A. Cortinovis, H. J. Ferreau, D. Lewandowski, and M. Mercangöz, "Safe and efficient operation of centrifugal compressors using linearized MPC," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 3982–3987.
- [2] R. Schulze, F. Dietel, J. Jäkel, and H. Richter, "Using an artificial immune system for classifying aerodynamic instabilities of centrifugal compressors.," in *NaBIC*, 2010, pp. 31–36.
- [3] Jayantha P. Liyanage and Uday Kumar, "Towards a value- based view on operations and maintenance performance management," *J. Qual. Maint. Eng.*, vol. 9, no. 4, pp. 333–350, Dec. 2003.
- [4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.
- [5] J. Lewitt, *Complete Guide to Predictive and Preventive Maintenance*. Industrial Press Inc., New York, 2002.
- [6] Z. Tian, T. Jin, B. Wu, and F. Ding, "Condition based maintenance optimization for wind power generation systems under continuous monitoring," *Renew. Energy*, vol. 36, no. 5, pp. 1502–1509, May 2011.
- [7] R. Bris, E. Châtelet, and F. Yalaoui, "New method to minimize the preventive maintenance cost of series-parallel systems," *Reliab. Eng. Syst. Saf.*, vol. 82, no. 3, pp. 247–255, Dec. 2003.
- [8] M. Samrout, F. Yalaoui, E. Châtelet, and N. Chebbo, "New methods to minimize the preventive maintenance cost of series-parallel systems using ant colony optimization," *Reliab. Eng. Syst. Saf.*, vol. 89, no. 3, pp. 346–354, Sep. 2005.
- [9] T. Xia, L. Xi, X. Zhou, and J. Lee, "Condition-based maintenance for intelligent monitored series system with independent machine failure modes," *Int. J. Prod. Res.*, vol. 51, no. 15, pp. 4585–4596, Aug. 2013.

- [10] “Anonymous, 2007, ‘Production Optimisation through Advanced Condition Monitoring of Upstream Oil and Gas Assets,’ Retrieved 23 April, 2010, from http://www.matrikon.com/portal/downloads/equipmentconditionmonitor/whitepaperecm_advancedconditionmonitoring.pdf.”
- [11] R. Ahmad and S. Kamaruddin, “An overview of time-based and condition-based maintenance in industrial application,” *Comput. Ind. Eng.*, vol. 63, no. 1, pp. 135–149, Aug. 2012.
- [12] J.-H. Shin and H.-B. Jun, “On condition based maintenance policy,” *J. Comput. Des. Eng.*, vol. 2, no. 2, pp. 119–127, Apr. 2015.
- [13] D. HALimi, A. Haifa, and E. BouAli, “Maintenance actions planning in industrial centrifugal compressor based on failure analysis,” *Eksplot. Niezawodn.*, vol. 16, no. 1, 2014.
- [14] T. Malmi and S. Ikäheimo, “Value Based Management practices—some evidence from the field,” *Manag. Account. Res.*, vol. 14, no. 3, pp. 235–254, Sep. 2003.
- [15] J. D. Martin and J. W. Petty, “Value Based Management: The Corporate Response to the Shareholder Revolution,” Oxford University Press, OUP Catalogue, 2001.
- [16] G.-L. Liao, “Optimal economic production quantity policy for randomly failing process with minimal repair, backorder and preventive maintenance,” *Int. J. Syst. Sci.*, vol. 44, no. 9, pp. 1602–1612, Sep. 2013.
- [17] R. Pascual, V. Meruane, and P. A. Rey, “On the effect of downtime costs and budget constraint on preventive and replacement policies,” *Reliab. Eng. Syst. Saf.*, vol. 93, no. 1, pp. 144–151, Jan. 2008.
- [18] N. Aretakis, I. Roumeliotis, G. Doumouras, and K. Mathioudakis, “Compressor washing economic analysis and optimization for power generation,” *Appl. Energy*, vol. 95, pp. 77–86, Jul. 2012.
- [19] G. M. Kopanos, D. P. Xenos, M. Ciccotti, E. N. Pistikopoulos, and N. F. Thornhill, “Optimization of a network of compressors in parallel: Operational and maintenance planning – The air separation plant case,” *Appl. Energy*, vol. 146, pp. 453–470, May 2015.
- [20] D. P. Xenos, G. M. Kopanos, M. Ciccotti, and N. F. Thornhill, “Operational optimization of networks of compressors considering condition-based maintenance,” *Comput. Chem. Eng.*, vol. 84, pp. 117–131, Jan. 2016.
- [21] R. Ahmad, S. Kamaruddin, M. Mokhtar, and I. P. Almanar, “Identifying the Best Fit Failure Distribution and the Parameters of Machine’s Component: A New Approach,” 2006.
- [22] M. Silverman, “Non-financial recognition: The most effective of rewards,” *Inst. Employ. Stud. Res. Netw.*, 2004.
- [23] M. Ciccotti, D. P. Xenos, A. E. Bouaswaig, N. F. Thornhill, and R. F. Martinez-Botas, “Online performance monitoring of industrial compressors using meanline modelling,” in *ASME Turbo Expo 2014: Turbine Technical Conference and Exposition*, 2014, p. V02DT42A003–V02DT42A003.
- [24] F. Milost, “Non-financial performance ratios as a management tool,” in *Proceedings of the 4th International Conference of the Faculty of Management Koper, University of Primorska*, 2003.
- [25] C. Jasch, “The use of Environmental Management Accounting (EMA) for identifying environmental costs,” *J. Clean. Prod.*, vol. 11, no. 6, pp. 667–676, Sep. 2003.
- [26] “Koul, Ashok. ‘Method and system for real-time prognosis analysis and usage based residual life assessment of turbine engine components and display.’ U.S. Patent No. 8,116,990. 14 Feb. 2012.”
- [27] P. D. T. O’Connor, P. O’Connor, and A. Kleyner, *Practical Reliability Engineering*. John Wiley & Sons, 2012.
- [28] J. Meyer and R. H. Rasche, *Kolmogorov-Smirnov Tests for Distribution Function Similarity with Applications to Portfolios of Common Stock*. National Bureau of Economic Research, 1989.
- [29] “Advanced Compressor Cleaning for Gas Turbines.pdf.” [Online]. Available: http://m.energy.siemens.com/hq/pool/hq/energy-topics/technical-papers/Advanced_Compressor_Cleaning_for_Gas_Turbines.pdf.

Session 4

Modelling, Analysis and Optimisation 2

Session Chair Prof. Yimin Shao

Planetary Gearbox Diagnosis Based on Morlet Wavelet Filter and Time Synchronous Averaging
(X. Xu, M. Zhao, J. Lin, C. Yan, J. Zhou, J. Wang)

Advanced Prognosis Method for Circuit Breaker Condition Monitoring
(N. Chen, R. Yu, J. Ruan, X. Zhang)

Application of Stochastic Resonance and Variational Mode Decomposition in Rolling Bearing Fault Diagnosis (J. Wang, S. Chen, C. Zhang, L. Deng, J. Hu)

Planetary Gearbox Diagnosis Based on Morlet Wavelet Filter and Time Synchronous Averaging

Xiaoqiang Xu¹, Ming Zhao¹, Jing Lin², Chang Yan¹, Jun Zhou³, Jianming Wang³

¹ Shaanxi Key Laboratory of Mechanical Product Quality Assurance and Diagnostics, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

² State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, China

³ Qinghai Huading Heavy Machine Tool CO., LTD, Xining 810000, China

ABSTRACT

Planetary gearboxes are widely used in automotive, aerospace and heavy industry due to its advantages of large transmission ratio and strong load-bearing capacity. The faults of planetary gearboxes may cause the machine to break down, resulting in huge economic loss and even catastrophic casualties. The prognostics and health management of planetary gearboxes has been attracting considerably increasing attention over the past decades. The frequency spectrum of the vibration signal can be very complicated even for healthy planetary gearboxes because of multiple gear meshes and multiple time-varying transfer paths, which makes the diagnosis of planetary gearbox very challenging. In this paper, a hybrid method based on optimal Morlet wavelet filter and time synchronous averaging (TSA) is presented to capture the fault-induced weak signature. With the proposed method, the complex Morlet wavelet filter is utilized to locate the fault-induced impulses in frequency domain. And the kurtosis criterion is employed to optimize the parameters of filter. Since the modulations of multiple time-varying transfer paths are dominating the signal at whole frequency range, the filtered signal is still blur because of the many unrelated components. Thus the TSA is further implemented to the envelope of filtered signal for de-noising and only the fault-induced transient components are preserved. Data from a planetary gearbox test rig is used to facilitate the comparison and evaluation of the methods. Two chip faults on a planetary gear are introduced by wire cutting. The results show that neither complex Morlet wavelet filter nor TSA alone can detect the faults directly because of the heavy noise. But the proposed hybrid method detects the chip faults successfully and the fault signature is recognizable in time domain.

Keywords: Morlet wavelet filter; time synchronous averaging (TSA); planetary gearbox; fault diagnosis; de-noising

Corresponding author: Jing Lin (email: jinglin@mail.xjtu.edu.cn)

1. INTRODUCTION

Planetary gearboxes are commonly used in many rotating machineries because of its advantages of compactness, efficiency and large power to weight ratios. Since their failures may lead to unscheduled maintenance or even large casualties, the early detection of the incipient fault is of vital importance. Vibration analysis has been the most extensively studied technique for planetary gearboxes diagnosis. The vibration signal of a planetary gearbox is complex because of the multiple, nonlinear gear meshes.[1] There are also multiple and time-varying transmission paths from the gear mesh points to the transducers mounted on the gearbox housing.[2] Thus the vibration signal of planetary gearbox is much more complicated than that of a fixed-axis gearbox. Various methods have been developed for the diagnosis of planetary gearboxes. Saxena [3] proposed a wavelet domain methodology for the analysis and feature extraction of the vibration data from a planetary gear system of the Blackhawk helicopter. Barszcz [4] applied spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Lei [5] proposed an adaptive stochastic resonance method for planetary gearbox fault diagnosis. Among all kinds of signal processing methods, the Morlet wavelet filtering [6] and TSA [7] have been demonstrated as effective and efficient

methods for weak signature enhancement working in two different ways. The Morlet wavelet filter essentially serves as a band-pass filter in frequency domain, and TSA serves as a comb filter to extract the interested harmonics. In this paper, the complex Morlet wavelet filter optimized by maximum kurtosis is utilized to extract the impulsive components. And the TSA is further implemented to the envelope of filtered signal for further de-noising. The proposed hybrid method is validated by the data from a planetary gearbox test rig.

2. THEORY

2.1. Morlet wavelet filter

The wavelet filter has been widely used to extract the fault-induced transients from the noisy vibration data. Generally, continuous wavelet transform (CWT) is preferable than discrete wavelet transform (DWT) for machinery fault diagnosis because the resolution is higher and it has more choices of wavelet base function. The wavelet base function normally need to be chosen according to the characteristic of practical signal. A typical fault-induced transient comprises an initial fast rise and a decaying part. The decaying part of the Morlet wavelet and that of the typical impact generated transient is similar, thus it is suggested to use the Morlet wavelet to construct the matching filter to extract the characteristic signal of the faults. [8] In this paper, the complex Morlet wavelet [6, 9] is employed and its formula is given as follows:

$$\psi(t) = \frac{\sigma}{\sqrt{\pi}} e^{-\sigma^2 t^2} e^{j2\pi f_0 t}. \quad (1)$$

The Fourier spectrum of the complex Morlet wavelet is

$$\Psi(f) = e^{-(\pi^2/\sigma^2)(f-f_0)^2}, \quad (2)$$

which has a Gaussian window shape in the frequency domain, where f_0 is the center frequency and σ is its bandwidth. The complex Morlet filter essentially is a band-pass filter within the frequency range of $[f_0 - \sigma/2, f_0 + \sigma/2]$. And the Morlet wavelet filter can be constructed as

$$WT(f_0, \sigma) = F^{-1}\{X(f)\Psi^*(f)\}. \quad (3)$$

In the implementation of Eq. (3), $X(f)$ is calculated only once, whereas $\Psi^*(f)$ is calculated for all possible combination of f_0 and σ . Therefore, the computation efficiency is relatively high. The maximum kurtosis criterion is utilized to optimized the parameters (f_0 and σ) of the filter.

2.2. Time synchronous averaging

The time synchronous averaging (TSA) is a common method for extracting a periodic component of interest from a noisy vibration signal and regarded as a powerful tool for the detection and diagnosis of planetary gear faults.[2] A deterministic signal can be extracted by TSA via enhancing the synchronous vibration components and reducing non-synchronous components and noise. Faults can often be detected in its early stages by direct inspection of the time averaged waveform. [7] And TSA has been widely used for vibration analysis of rotating machine components. [10]

With the interested period determined, we divide the time-domain waveform into segments in length of period we concern, and then average the corresponding data among all segments. In such direct averaging method, the period actually used is slightly different from real interested period and possess period cutting

error introduced by discrete sampling. Liu [11] introduce an improved algorithm for the TSA which basically solves the problem caused by the period cutting error. Suppose that there is a discrete time series with sample interval Δt :

$$x(n), n=0,1,\dots,N-1, \quad (4)$$

and the interested component of period T is to be separated out; then algorithm for the TSA proposed by Liu [11] is

$$y(n) = \frac{1}{K} \sum_{i=0}^{K-1} x(n - m_i), n = m_{k-1}, \dots, N-1 \quad (5)$$

where $y(n)$ is the component of interest, K is the number of averaging periods, $m_i = \text{round}(iT/\Delta t)$. The algorithm is straightforward and easy implemented. It employed in this paper to implement the TSA.

2.3. The proposed hybrid method

Due to the heavy noise of vibration signal measured from planetary gearbox, Morlet wavelet filter and TSA alone normally cannot achieve a satisfactory de-noising result. The hybrid method as shown in figure 1 is proposed to combine the two techniques and extract fault-induced transient component. First, the Morlet filter with maximum kurtosis is used to find the impulsive components in the measured signal. And then the TSA is implemented to the envelope of the filtered signal (wavelet coefficients) to further recover the interested vibration component. According to the waveform derived by the proposed method, the incipient fault of the planetary gearbox can be identified.

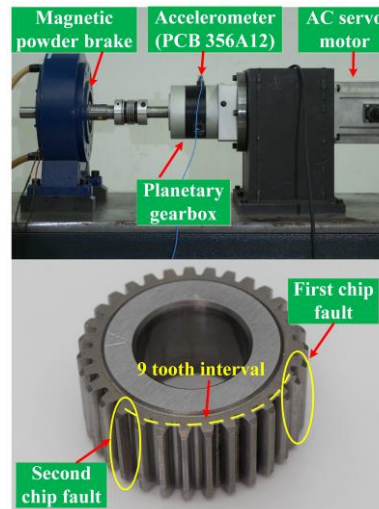
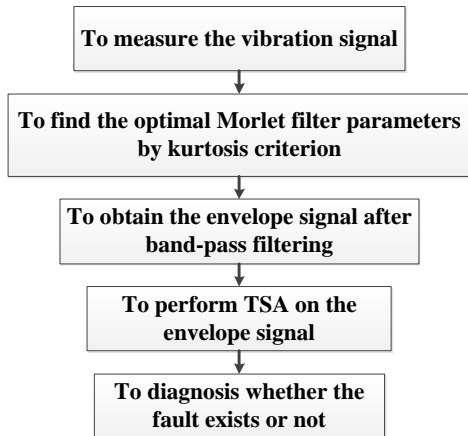


Figure 1. The flow chart of the proposed method.

Figure 2. The planetary gearbox test rig and the fault planetary gear.

3. EXPERIMENT

Data of a planetary gearbox test rig is utilized to illustrate the effectiveness of the proposed method. The test rig consists of an AC servo motor, a single stage planetary gearbox and a magnetic powder brake as shown in figure 2 (a). Two chip faults with 9 tooth interval are introduced by wire cutting in a planetary gear as shown in figure 2 (b). The structure parameters of the planetary gearbox are given in table 1, and the fault characteristic frequency of each component is calculated in table 2. The rotating speed of the output shaft is measured by the tachometer. An ICP accelerometer PCB 356A12 with sensitivity of 100mV/g is mounted on the ring of planetary gearbox to measure the vibration signal. The vertical vibration with 4

seconds length sampled at 20000 Hz is given in figure 3 (b). The kurtosis distribution for different parameters of Morlet wavelet filter is given in figure 3 (a), from which the optimal filter ($f_0=7000\text{Hz}$, $\sigma=4000\text{Hz}$) with maximum kurtosis is determined. The optimal band-pass filter is presented in figure 3 (c) and the filtered signal is shown in figure 4 (a). Periodic transient signature cannot be observed from the waveform. We further implemented the TSA to the envelope of the filtered signal to extract the planetary gear related vibration. Figure 4 (c) shows the result of proposed method, the two faults and their relative position are clearly identified. We also apply the TSA directly to the original signal to extract the planetary gear related components and the result is given in figure 4 (b). Only one fault per planetary gear rotation can be blurry observed. In this case, neither Morlet wavelet filter nor TSA alone can detect the faults directly because of the heavy noise. But the proposed hybrid method detects the chip faults successfully and the fault signature is recognizable in time domain.

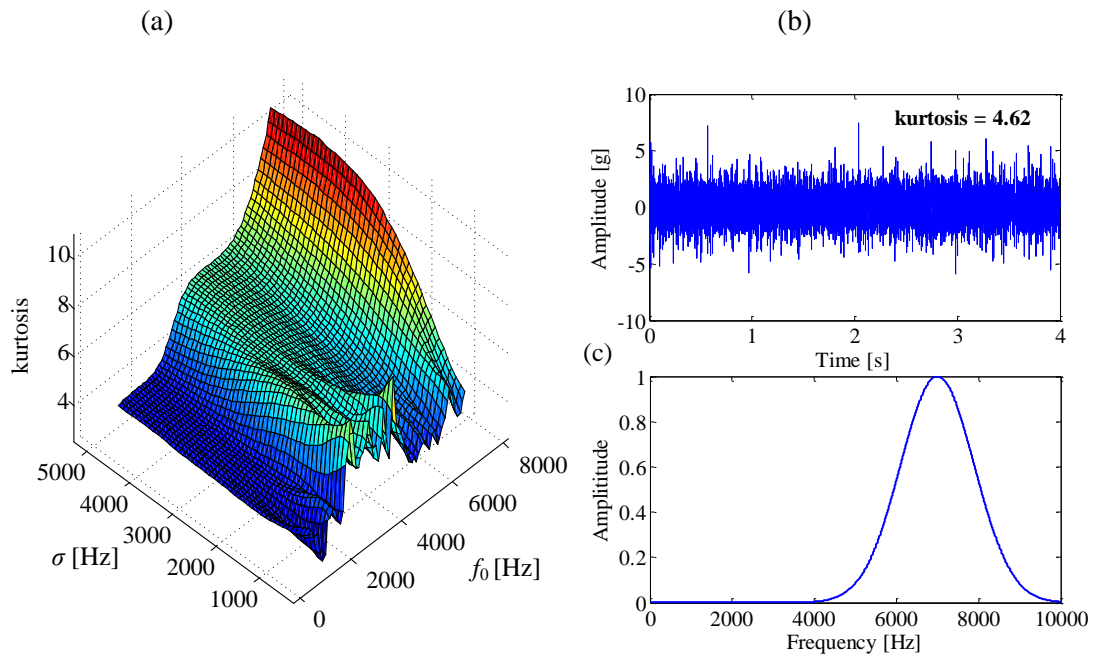


Figure 3. (a) The kurtosis distribution for different parameters of Morlet wavelet filter; (b) the original vibration waveform; (c) the optimal filter with band-pass frequency from 5000Hz to 9000Hz.

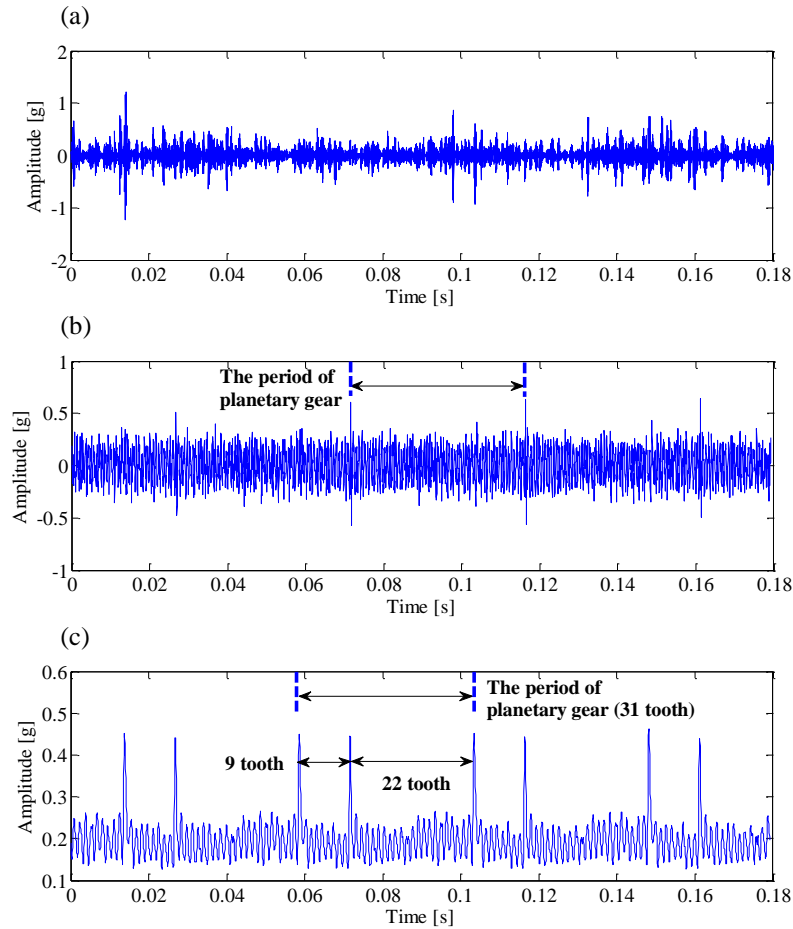


Figure 4. The results for the vibration signal of a planetary gear with two chip faults using: (a) Morlet wavelet filter; (b) TSA; (c) the proposed hybrid method.

4. CONCLUSION

In this paper, a hybrid method combined Morlet wavelet filter and time synchronous averaging is proposed. It improves the SNR of the original signal significantly and is especially suitable for planetary gearbox diagnosis. The Morlet wavelet filter serves as a band-pass filter to extract the impulsive components in frequency domain. And the TSA is further implemented to the envelope of filtered signal for de-noising and excluding other vibration components. The proposed method is demonstrated by data collected from a real planetary gearbox test rig. The proposed method successfully extracts the fault-induced transients and none of the two original techniques can detect the faults directly.

5. ACKNOWLEDGEMENT

This project is supported by the National Science and Technology Major Project (2014ZX04001051), which is highly appreciated by the authors.

REFERENCE

- [1] M. Botman, Vibration Measurements on Planetary Gears of Aircraft Turbine Engines, *Journal of Aircraft*, 17 (1980) 351-357.
- [2] D.M. Blunt, J.A. Keller, Detection of a fatigue crack in a UH-60A planet gear carrier using vibration analysis, *Mechanical Systems and Signal Processing*, 20 (2006) 2095-2111.

- [3] A. Saxena, W. Biqing, G. Vachtsevanos, A methodology for analyzing vibration data from planetary gear systems using complex Morlet wavelets, American Control Conference, 2005. Proceedings of the 2005, 2005, pp. 4730-4735 vol. 4737.
- [4] T. Barszcz, R.B. Randall, Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mechanical Systems and Signal Processing, 23 (2009) 1352-1365.
- [5] Y. Lei, D. Han, J. Lin, Z. He, Planetary gearbox fault diagnosis using an adaptive stochastic resonance method, Mechanical Systems and Signal Processing, 38 (2013) 113-124.
- [6] W. Su, F. Wang, H. Zhu, Z. Zhang, Z. Guo, Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement, Mechanical Systems and Signal Processing, 24 (2010) 1458-1472.
- [7] P.D. McFadden, Interpolation techniques for time domain averaging of gear vibration, Mechanical Systems and Signal Processing, 3 (1989) 87-97.
- [8] W. Yang, X. Ren, Detecting impulses in mechanical signals by wavelets, EURASIP Journal on Applied Signal Processing, 8 (2004) 1156-1162.
- [9] N.G. Nikolaou, I.A. Antoniadis, Demodulation of Vibration Signals Generated by Defects in Rolling Element Bearings Using Complex Shifted Morlet Wavelets, Mechanical Systems and Signal Processing, 16 (2002) 677-694.
- [10] M. Zhao, J. Lin, Y. Lei, X. Wang, Flexible time domain averaging technique, Chinese Journal of Mechanical Engineering, 26 (2013) 1022-1030.
- [11] L. Hongxing, Z. Hongfu, J. Chengyu, Q. Liangsheng, An Improved Algorithm for Direct Time-Domain Averaging, Mechanical Systems and Signal Processing, 14 (2000) 279-285.

Advanced Prognosis Method for Circuit Breaker Condition Monitoring

Niya Chen¹, RongRong Yu¹, JiaYang Ruan¹, Xin Zhang²

¹ ABB Corporate Research Center, 100015, Beijing, China

² ABB Electrification Product Medium Voltage, Technology Center

ABSTRACT

Circuit breaker is one of the most critical equipment in substation, whose safety and reliability are of high importance to the overall power grid. In this paper, a prognosis method based on pressure sensor signal is proposed to estimate the condition of circuit breaker.

The proposed method is able to process complex pressure signal which has abundant frequency components, by extracting its key features using wavelet decomposition. Afterwards the method evaluates the dissimilarity of different samples based on quantifying the difference between their feature vectors. By comparing test sample with normal training samples, the dissimilarity can indicate if the test circuit breaker is normal or defective.

Data from medium voltage (40.5 kV) circuit breaker are used to train and validate the proposed method. In the experiment, about 1000 normal samples and 26 defective samples (2 kinds of defects) are used for model validation. The results show that the built model can reflect condition of circuit breaker accurately, with a 90% detection rate, and almost 0% false alarm rate.

Keywords: Prognosis, circuit breaker, condition monitoring, wavelet transform

Corresponding author: Niya Chen (email: niya.chen@cn.abb.com)

1. INTRODUCTION

Circuit breaker (CB) plays a very important role in the power system. CBs need to be reliable since their failure can cause major issues with power system protection and control [1]. Therefore, normally many sensors are installed on CB to monitor its operation status [2]. As a matter of fact, in China, certain online monitoring systems are required or preferred for CBs by State Grid Corporation of China (SGCC) in its cooperation standard [3]. However, nowadays the massive data from sensors are usually directly shown to the user, thus is difficult to understand or use. Therefore, it's worthwhile developing CB prognosis method, which can clearly indicate the health status of CB and predict failure which enables predictive maintenance [4][5].

Throughout the years, many kinds of circuit breaker online monitoring system have been developed and the corresponding diagnosis/prognosis methods have been studied. Guan [5] investigated the correlation between coil current and circuit breaker mechanical status, so that the coil current data can be used to diagnose the mechanical failure in CB. Straightforward time-domain calculation is used in the paper for coil current curve analysis. Rudd [6] proposed a circuit breaker prognostics system, which ranks circuit breakers in order of maintenance priority, by monitoring and analyzing the SF6 density in CBs. Mechanical vibration signals also have been used to detect mechanical fault in circuit breaker, and different kinds of methods have been used, such as wavelet packet analysis [7] and artificial neural network [8].

Pressure sensor installed in CB can measure the pressing force of contact in closing operation, hence can provide a clear indication of CB status. However, the pressure signal usually contains abundant frequency components thus can hardly be analyzed through classical time-domain method. In this paper, a new prognosis method is proposed to estimate the CB status using pressure signal. The method first extract key

features of pressure signal using wavelet decomposition, then calculate the dissimilarity between different samples. The CB status can be estimated through quantifying the dissimilarity between online monitored test sample and pre-stored normal training samples.

The paper is organized as following. Section 2 describes the structure of the proposed prognosis method, including both the training process and testing steps. Section 3 introduces the feature extraction method used in the proposed model, while Section 4 illustrates how to determine the status of circuit breaker by quantifying dissimilarity of test sample with normal samples. The information of dataset, the steps for data pre-processing, the detailed model validation process, and the experimental results are presented in Section 5. Finally Section 6 draws the overall conclusions.

2. MODELING PROCESS OF PROGNOSIS METHOD

The basic assumption in the proposed prognosis method is that sensor signals measured in normal status shall be similar to each other. In other words, if test data found as relatively different from normal training data, it is indicated that some abnormality appears in Circuit Breaker. Therefore, the key problem here is how to design a scheme to reasonably measure the similarity/dissimilarity among sensor signals.

To solve the problem, in our prognosis method, there are three main parts: 1) extract feature of the monitored signal(s) due to the spectrum analysis; 2) construct benchmark matrix by the features of partial normal training samples, and calculate the dissimilarity of the rest normal training samples to the benchmark matrix; 3) determine threshold according to the dissimilarity variation of normal training samples, making sure that the threshold is reasonably larger than the normal dissimilarity. As shown in figure 1, from training set, 2 modules are built and shall be used for test data – the normal feature matrix and threshold. Details of the modeling process are illustrated in figure 1 and described below.

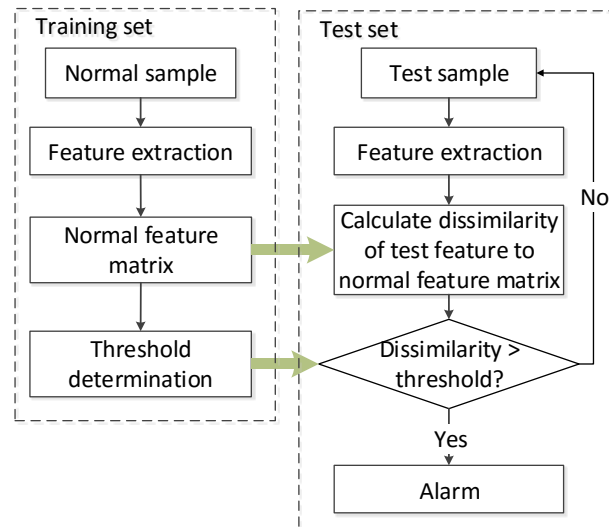


Figure 1 Modelling process of the proposed prognosis method

Firstly the dataset used in this paper are divided into 2 parts: training set for model training, test set for model performance validation. Note that the training set only needs the data monitored when CB is in normal condition, while test set contains both normal and defective CB data.

Usually 1 train/test sample means the sensor data measured in 1 CB operation, hence is a curve contains possibly thousands of points. Therefore, key features need to be extracted from original curve to compress the data quantity. Detailed feature extraction method is given in Section 3.

Secondly, part of the normal data in training set is used to construct a ‘benchmark matrix’. Thirdly, dissimilarity between the rest normal training samples and the benchmark matrix are calculated. Intuitively and also classically in math, one way to calculate the dissimilarity is to calculate the ‘distance’ between the

sample and the benchmark matrix in feature space. Since the matrix only includes features of normal CB data, the dissimilarity ('distance' between the benchmark matrix and the test sample) is very small if the CB is still normal, likewise, the dissimilarity value shall be large when CB is abnormal. The specific equation for dissimilarity/distance calculation will be given in Section 4. Finally, in the training set, now all the dissimilarity data are known for normal samples. A threshold value can be decided to guarantee all the normal dissimilarities are much lower than the threshold.

From the training set, a benchmark matrix and a threshold value that can separate abnormal and normal sample are obtained. Therefore, with any new input from the test set, to determine the current CB status, the process should be: firstly, key features are extracted from the test sample; secondly, the dissimilarity can be obtained by calculating the 'distance' between the test sample and benchmark matrix; finally, compare the residual with the threshold value, if it is smaller than the threshold, the CB can be determined as normal, otherwise if it is larger, then it can be revealed that there is probably a potential failure going on in this CB.

3. FEATURE EXTRACTION METHOD

Feature extraction method for pressure signal is introduced in this section. Pressure sensor can be installed in different locations in circuit breaker, and the measured signal often contains abundant frequency components. In this paper, pull rod force signal is analyzed as an example, to illustrate how key features are extracted for pressure signals.

3.1. Necessity and effectiveness of wavelet decomposition

Pull rod force in circuit breaker, which is measured by pressure sensor, has vibrating shape caused by redundant frequency components. Dozens of monitored pull rod force curves under CB normal and defective condition are shown in figure 2.

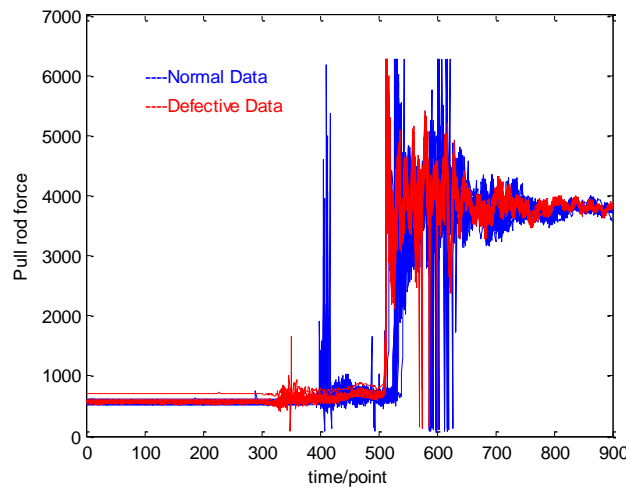


Figure 2 pull rod force curve under normal / failed condition

There are 36 Normal samples and 26 defective samples plotted together in figure 2, with the normal samples plotted in blue line, and defective ones in red line. It can be seen that there indeed are some differences between normal and defective samples, but these differences can hardly be represented by straightforward features through direct time-domain calculation.

A reasonable assumption is that the main tendency in defective pull rod force curve shall be different from normal ones. Therefore, wavelet transform is used here to extract the major trend, i.e. low frequency part in pull rod force curve. The extracted low frequency curves are shown in figure 3. Detailed principle and equations of wavelet transform are given in section 3.3.

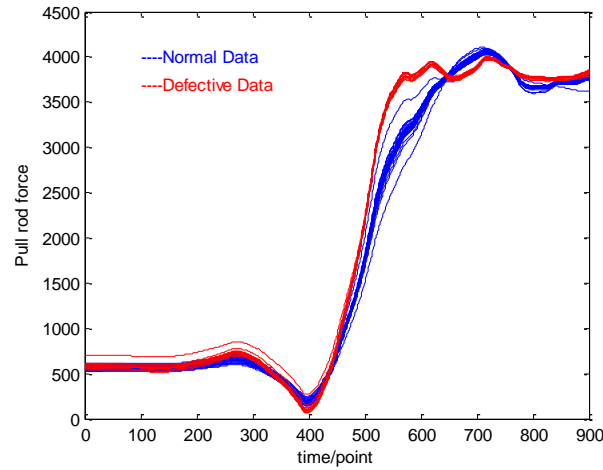


Figure 3 Decomposed pull rod force curves under normal / failed condition

To be more specific, 7 level-‘db4’ wavelet decomposition method is chosen in this paper to process the monitored pull rod force curves, to better distinguish the normal and defective samples. The effectiveness of wavelet decomposition in failure detection can easily be proved, since in the decomposed curves, the normal and defective curves are obvious different from each other, hence can be distinguished easily by prognosis model. Furthermore, feature extraction of the decomposed curves can be done in a simple way as presented in section 3.2.

3.2. Feature extraction of decomposed curve

The original pull rod force curve from closing operation is plotted in the left part below, which needs to be decomposed by wavelet transformation first. The decomposed low frequency sub-series is plotted in the right part of figure 4 below, and now features can be easily extracted.

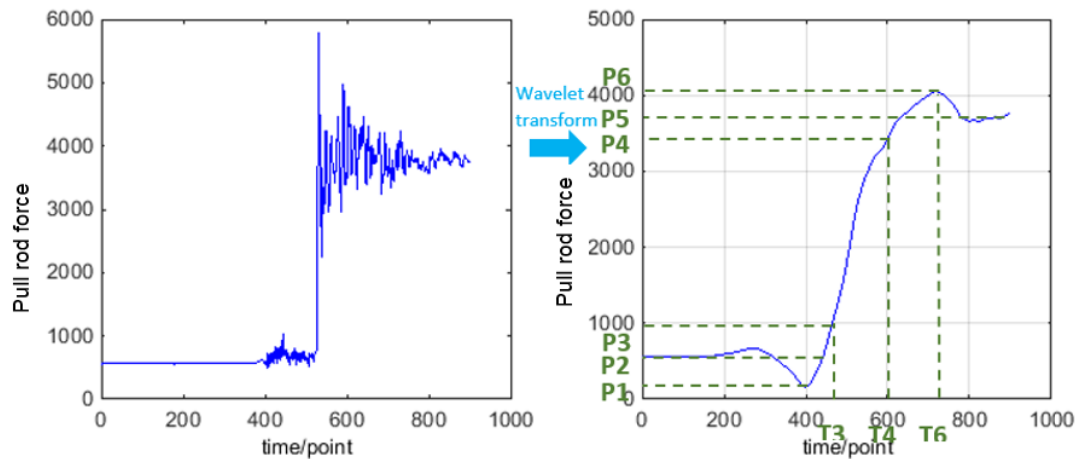


Figure 4 Feature calculation for pull rod force

The low frequency component of pull rod force curve has a clear shape, and key features can be extracted according to the curve shape. The equations for extracting key features are:

$$F_1 = (P_4 - P_3) / (T_4 - T_3)$$

$$F_2 = P_6 - P_1$$

$$F_3 = T_4 - T_3 \quad (1)$$

$$F_4 = P_5 - P_2$$

$$F_5 = P_6 - P_5$$

where P_1 is the minimum value of travel curve, P_2 is the value of starting point, P_5 is the value of curve end, and P_6 is the maximum value. P_3 and P_4 are located by the following equations:

$$P_3 = P_2 + (P_5 - P_2) \cdot 25\% \quad (2)$$

$$P_4 = P_2 + (P_5 - P_2) \cdot 85\% \quad (3)$$

Finally the original curve can be represented by this feature vector which only contains 5 scalars.

3.3. Wavelet transformation

The wavelet method is used in this paper to decompose the original monitored curve of complex signal curve in one operation, into a set of sub-series. With the filtering effect of the wavelet decomposition, these sub-series present different frequency components of the original monitored complex curve, and therefore can be analyzed more clearly with the main trend extracted.

Wavelet method can be divided in two categories: continuous wavelet transform (CWT) and discrete wavelet transform (DWT), the details can be found in [10][11]. Usually DWT is more realistic and convenience to realize in computer program, and DWT is the method used in this paper. Here the core equations of CWT and DWT are briefly introduced.

In CWT, a wavelet can be defined as function $\Psi(t)$ with a zero mean.

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (4)$$

Wavelet Transform (WT) is like Fourier transform, the core function of which is to analyze the spectral components in signal. The advantage of WT is that it has balanced resolution in both time domain and frequency domain. Briefly speaking, the process of CWT can be seen as calculating the ‘correspondency’ of signal $f(t)$ to wavelets with different scales a and translation b ,

$$wf(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (5)$$

Since usually the signal restored in computer is a discrete time function (an array), DWT is more likely used in program. The discrete signal is denoted as $x[n]$ hereinafter.

Differ from CWT, the DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low-pass and high-pass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time domain signal:

$$y_{high}[k] = \sum_n x[n] \cdot g[2k - n] \quad (6)$$

$$y_{low}[k] = \sum_n x[n] \cdot h[2k - n] \quad (7)$$

where $g[n]$ is a high pass filter, $h[n]$ is a low pass filter, and $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high-pass and low-pass filters, respectively.

The above procedure can be repeated for further decomposition, so that the original signal can be decomposed into multiple frequency components in narrower range.

4. DISSIMILARITY CALCULATION AND THRESHOLD DETERMINATION

Euclidean distance [12] is selected to calculate the dissimilarity between 2 features in this paper. Denote an offline normal dataset of target CB as $\{C_1, C_2, \dots, C_n\}$, while each C_i is one pull rod force curve monitored in one operation. According to feature extraction method, several features can be extracted from the curve data. Therefore, the monitored data C_t can be transferred to X_t which includes 5 scalar. Naturally, the offline normal dataset $\{C_1, C_2, \dots, C_n\}$ becomes $\{X_1, X_2, \dots, X_n\}$.

As described in Section 2, firstly the offline normal dataset is split evenly into 2 parts D and D_* , while D includes features of some normal samples, and D_* includes features of the rest of the samples. D is considered as a benchmark matrix for residual calculation, and D_* will be used for threshold determination.

$$D = \{X_1, X_2, \dots, X_m\} \quad (8)$$

$$D_* = \{X_{m+1}, X_{m+2}, \dots, X_n\} \quad (9)$$

Then the dissimilarity of each test data X_t can be calculated as:

$$S_t = \text{mean}(D^T \otimes X_t) \quad (10)$$

$$D^T \otimes X_t = \begin{bmatrix} X_1^T \otimes X_t \\ \vdots \\ X_m^T \otimes X_t \end{bmatrix} \quad (11a)$$

$$X \otimes Y = \sqrt{\sum_{i=1}^t (x_i - y_i)^2} \quad (11b)$$

where S_t is the calculated dissimilarity.

Calculate the dissimilarity set $S_{train} = [S_{m+1}, S_{m+2}, \dots, S_n]$ of all the feature vectors in D_* , according to equation (10). Threshold value is determined as,

$$Thre = N \cdot \max(S_{train}) \quad (12)$$

Where N is set to 4 in experiments in this paper.

With dissimilarity defined and threshold determined, status determination is quite straight forward – for any measured data X_t , the corresponding status is abnormal if $S_t > Thre$, otherwise is normal.

In conclusion, $SS_t = S_t - Thre$ can be shown to customer indicating the health condition of CB, where $SS_t \leq 0$ means normal, and $SS_t > 0$ means abnormal, i.e. defect developing in CB.

5. EXPERIMENTAL RESULTS OF PROGNOSIS MODEL

5.1. Experimental database

Several samples are collected from MV CB, to train and validate our prognosis model. The detailed information are given in the table below.

Table 1: Experimental database information

| CB Type | Voltage level | Normal data | Failed data | Defective component |
|-----------|---------------|-------------|-------------|---------------------|
| Vacuum CB | 40.5 kV | 1062 | 26 | Spring, damper |

As shown in table 1, data are collected from a 40.5 kV vacuum Circuit Breaker, including 1062 normal samples and 26 defective samples. The defective samples are collected from two kinds of manual defects: the spring defect and damper defect.

5.2. Data pre-processing

In the original dataset, there are some errors caused by wrong communication or sensor failure. Therefore, actions for error elimination are necessary. The original data is processed as follows.

1. Delete error communication points. When the communication in monitoring system appears error, the received data point is zero or 'NaN'. Therefore, the data points recorded as 'NaN' or unreasonable zeroes are deleted.
2. Delete obviously unreasonable data. Failed sensor constantly sends back obvious physically unreasonable data, like pressure less than 0 bar, etc. All the samples with unreasonable values are deleted.
3. Delete unreasonable peak points. Occasional disturbance on sensor might lead to peak value in measurement – peak value means one extremely high or low value comparing to previous and next measurement, which is impossible due to physical principle.

5.3. Prognosis results

In the experiment, 100 normal samples selected as the training set, the rest of the normal samples and all the 26 defective samples are used as the testing set. Partial prognosis results are presented in figure 5.

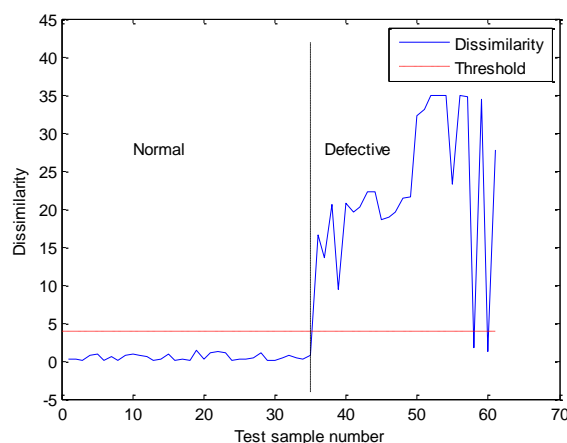


Figure 5 Residual error by pull rod force

It is clearly seen from figure 5 that the dissimilarity of normal samples are in a small range far less than the threshold, while the dissimilarity of defective samples increase obviously and mostly passes threshold line. The distinction degree is quite high considering the large gap between normal and abnormal results.

Note that some points in the curve end obviously decreases, of which 2 points are even below the threshold line. That's because the defective samples used in this paper are not from one developing defect continuously, but from 2 different kinds of independent manual defects. The dissimilarity can be different depending on different kinds of defects.

Table 2: Experimental results

| Test data type | Number of samples | True prognosis | Statistics |
|----------------|-------------------|----------------|----------------------|
| Normal | 962 | 962 | False alarm: ~0% |
| Defective data | 26 | 24 | Detection rate: ~90% |

The experimental results show that the prognosis model can stably detect the defects while avoiding false alarm. The statistical results proved that the prognosis model has an around 90% detection rate, and $< 1/963$ false alarm rate.

6. CONCLUSION

Prognosis model for Circuit Breaker status monitoring is proposed in this paper, based on wavelet and dissimilarity calculation method. The model is trained and validated using MV CB data.

To process the original monitored curves into features that are suitable for modeling, wavelet decomposition method is used to deal with complex pressure signal which has abundant components in different frequency range. Dissimilarity between test feature vectors and normal feature benchmark matrix are calculated using Euclidean distance, and is compared to pre-defined threshold to determine the CB health status.

The experimental results show that the prognosis model can stably detect defects and avoid false alarm. The statistical results proved that the prognosis model has an around 90% detection rate, and $< 1/963$ false alarm rate.

REFERENCES

- [1] Anton Janssen, Dirk Makareinis, Carl-Ejnar Sölver. (2014) International Surveys on Circuit-Breaker Reliability Data for Substation and System Studies, IEEE Trans on Power Delivery, vol. 29, No. 2, April.
- [2] Michael Stanek. (2000) Model-aided diagnosis for high voltage circuit breakers, Swiss Federal Institute of Technology Zurich, PhD dissertation, p63.
- [3] SGCC. (2010) Technical guide for on-line monitoring system of transformation equipment, Q/GDW 534 - 2010, p3.
- [4] SS Biswas, AK Srivastava, D Whitehead. (2014) A Real Time Data-Driven Algorithm for Health Diagnosis and Prognosis of a Circuit Breaker Trip Assembly, IEEE Transactions on Industrial Electronics, 62(6):1-1.
- [5] Tat Wai (Alan) Leung. (2007) Predictive Maintenance of Circuit Breakers, University of Waterloo, PhD dissertation.
- [6] Yonggang Guan, Xinxia Peng, Beiyang Liu, et al. (2014) The correlation between the coil current and the operation process of a high voltage circuit breaker. International Conference on Condition Monitoring and diagnosis 2014. Jeju, Korea, Sept. 21-25, 2014.
- [7] S. E. Rudd, V. M. Catterson, et al. (2011) Circuit breaker prognostics using SF6 data. IEEE power and energy society general meeting. IEEE Power and Energy Society General Meeting PESGM. New York, pp. 1-6.

- [8] YonPeng Meng, Shenli Jia and Mingzhe Rong. (2004) Condition monitoring of vacuum circuit breakers using vibration analysis, XXIst International Symposium on Discharges and Electrical Insulation in Vacuum, 2004. Proceedings. ISDEIV., Yalta, Crimea, pp. 341-344.
- [9] Yongpeng Meng, Shenli Jia, Mingzhe Rong, (2005) Mechanical condition monitoring of vacuum circuit breakers using artificial neural network, Article in IEICE Transactions on Electronics, August.
- [10] Smola, A. J., Scholkopf, B. (1998) A tutorial on support vector regression. *Statistics and Computing*, 14(3), pp. 199–222.
- [11] Hang T. Nguyen, Ian T. Nabney (2010) Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy*, 35(9), pp. 3674-3685.
- [12] Deza, Elena; Deza, Michel Marie (2009). *Encyclopedia of Distances*. Springer. p. 94.

Application of Stochastic Resonance and Variational Mode Decomposition in Rolling Bearing Fault Diagnosis

Jianguo Wang¹, Shuai Chen^{1*}, Chao Zhang¹, Liming Deng², Juan Hu²

¹ School of Mechanical engineering, Inner Mongolia University of Science and Technology, Baotou, 014010, China

² Chongqing Pure-Smart Science & Technology CO. LTD, Chongqing, 400039, China

ABSTRACT

Aiming at the problem that noise in the bearing fault signal will reduce the readability of the signal, a fault diagnosis method, based on stochastic resonance (SR) and variational mode decomposition (VMD), is proposed in this paper. Firstly, stochastic resonance technology is used to process the bearing fault signal; Secondly, original vibration signal is decomposed into a finite number of different frequency band stationary intrinsic mode functions (IMF) using the variational mode decomposition method; Finally, each IMF component is analyzed to find the fault frequency by the method. Compared with the traditional denoising method, stochastic resonance can enhance weak cycle impact composition in the hidden of the fault signal. The results show that the proposed method can extract the bearing fault characteristic frequency effectively.

Keywords: stochastic resonance ; variational mode decomposition ; intrinsic mode functions; fault characteristic frequency; fault diagnosis.

Corresponding author: Chen Shuai (email: 1054543069@qq.com)

1. INTRODUCTION

Large-scale machinery equipment usually work under severe working conditions, in which faults will occur inevitably. The failure of mechanical equipment will cause serious economic losses, while the bearing in mechanical equipments will be easily damaged. The bearing faults seriously affect the safe operation of equipment, so it is necessary to analyze vibration signals of bearings and to detect the damage location and degree [1]. Because the transmission path of vibration signal is complex and changeable, the fault signals picked up by sensors are nonlinear and non-stationary, which brings difficulties to the fault diagnosis [2].

In order to highlight the weak impact component in the fault signal, the stochastic resonance technique [3] is introduced and widely used in the feature extraction of fault signals. Different from the traditional noise reduction methods, stochastic resonance is a method to enhance the weak components of bearing faults by using noise. The reference [4] presents a cascade of stochastic resonance method in feature extraction for gear shock signal. The reference [5] puts forward a kind of adaptive stochastic resonance method, in which the signal-to-noise ratio could be used as the performance evaluation index of stochastic resonance system. The stochastic resonance method can only deal with the signal with small parameters (amplitude, frequency, noise intensity). The reference [6] presents a variable scale method.

The variational mode decomposition method (VMD) [7] is a non-recursive decomposition method, in which the central frequency and bandwidth of the intrinsic mode function are updated to achieve adaptive decomposition. Compared with empirical mode decomposition (EMD) and local mean decomposition (LMD), VMD is essentially a set of adaptive Wiener filters, so there are no aliasing and end effects due to the envelope problem and recursive operation. In reference [8], a method, based on VMD and approximate entropy, was used to process the vibration signals of hydroelectric generating units. Reference [9] presents the multi kernel support vector machine combined with VMD. Through the calculation of the time domain characteristics of each IMF component, the pattern recognition of rolling bearing fault can be detected. In

reference [10], aiming at the problem of parameter selection in VMD, an adaptive VMD method based on artificial fish swarm is proposed. The experimental results show the effectiveness of the method.

However, a single signal processing method is not enough to extract the fault characteristic frequency. Based on this, the method based on the combination of stochastic resonance and VMD is proposed in this paper. Firstly, the stochastic resonance method was used to denoise the vibration signal of fault bearing. Secondly, the signal is decomposed by the VMD method. Finally, in order to find the fault frequency of bearing, the spectrum was used to demodulate the intrinsic mode function (IMF) components.

2. STOCHASTIC RESONANCE

The weak periodic impact component is enhanced by adding the noise to stochastic resonance system. The research is based on classical bistable model, and can be represented with nonlinear langevin equation [4]

$$\begin{aligned}\frac{dx}{dt} &= -\frac{dU}{dx} + s(t) + n(t) \\ n(t) &= \sqrt{2D}\xi(t)\end{aligned}\quad (1)$$

where $x(t)$ is the output signal of the system, $s(t)$ is the input signal of the system, $\xi(t)$ is a gaussian white noise, and its mean value is 0, the variance is 1, D is the noise intensity. $U(X)$ is the potential function of bistable system, which is defined as follows:

$$U(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4 \quad (2)$$

The nonlinear langevin equation can be expressed as follows:

$$\frac{dx}{dt} = ax - bx^3 + s(t) + n(t) \quad (3)$$

The system parameters a , b is the set of real Numbers, according to the formula (2), Potential function $U(X)$ has two steady state points $x_{\pm} = \pm\sqrt{a/b}$ and one unsteady state point $x=0$, The potential barrier height $\Delta U = a^2/4b$. According to the nonlinear langevin equation, In fact, the output of the system $x(t)$ is a trajectory that the input signal $s(t)$ and noise $n(t)$ under the synergy of Brownian particles. When only the existence of periodic input signal $s(t)$, brown particle just move in a potential well, not jump into another potential well. When the periodic input signal $s(t)$ and noise $n(t)$ exist at the same time, the brownian particles can accumulates enough energy jump back and forth in the two potential well, and produce the same frequency vibration as the input signal that is the stochastic resonance, part of the noise energy transform into useful signal energy by this time, thus output signal-to-noise ratio(SNR) of the stochastic resonance system is greatly improved, and which effectively extract the weak signal from strong background noise.

But general stochastic resonance system is limited by a small parameters, namely amplitude, frequency of the input signal $s(t)$ and noise intensity is far less than 1. But the bearing signals are mostly larger parameters in the engineering practice, which cannot use stochastic resonance of small parameters to deal with the signal of the large parameters in the fault diagnosis. So this article use variable scale signal processing method[6] that it has been widely used in large signal processing parameters to compress the bearing signal, which make it meet the requirements of small parameter stochastic resonance. Through this transformation, each frequency component of the original engineer signal that is relative to the sampling frequency of new definition for the normalized again, so it does not change any nature of original measured signal. The principle of scale transformation as follows:

- (1) Setting a frequency compression ratio R ;
- (2) According to the R defines a compressed sampling frequency $f_{sr}=f_s/R$.
- (3) Setting the step length of Runge kutta method $h=1/f_{sr}$.

Constructing a large parameters signal that is $s(t)=0.5\times\cos(2\pi\times30t)$, added white noise and the intensity is nine, the sample frequency $f_s=2000$ and sampling points $n=2000$. Time domain and frequency domain waveform of the simulation signal is shown in figures 1 and 2. Fault characteristics of signal isn't found in time domain waveform, and due to strong noise interfered, frequency characteristic Spectral lines is also difficult to identify in the frequency spectrum.

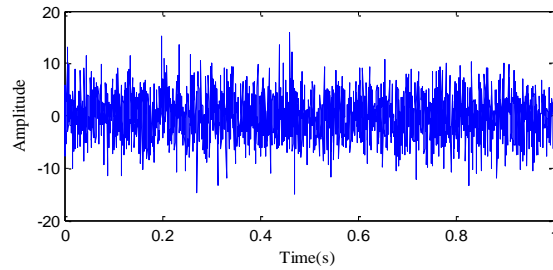


Figure.1 The time domain waveform of simulation signal

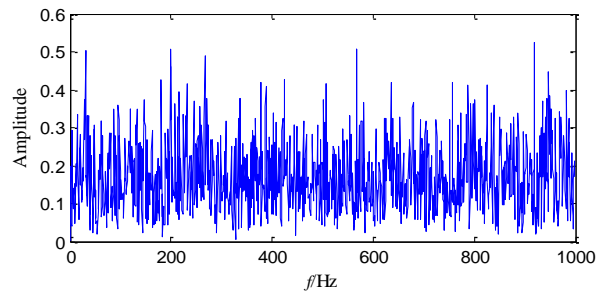


Figure.2 The spectrum of simulation signal

Setting frequency compression ratio is 500 and $a, b=1$. Time and frequency domain waveform of the output signal of stochastic resonance is shown in figure3 and figure 4. The noise can be restrained, which can be seen from figure 3. In figure 4, we can find the frequency $f_0=30\text{Hz}$ in the output signal spectrum of stochastic resonance.

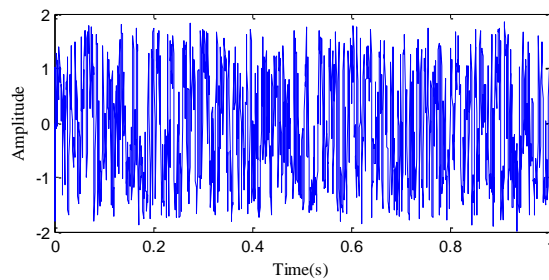


Figure.3 The output signal time domain waveform of stochastic resonance

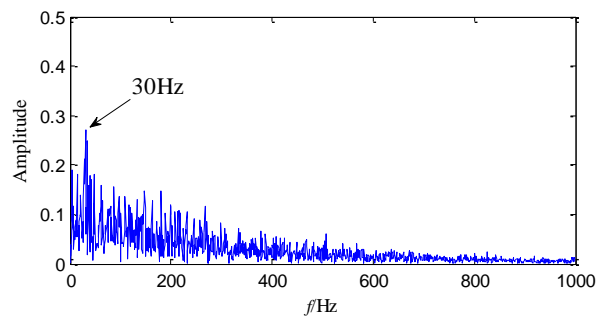


Figure.4 The output signal spectrum of stochastic resonance

3. VARIATIONAL MODE DECOMPOSITION

VMD, a new signal Decomposition method, is put forward in 2014 by Dragomiretskiy K and Zosso D. By solving constrained variational problem, arbitrarily complex multi-component signal will be adaptive decomposed into a series of the IMF components based on center frequency and bandwidth limited. In the process of solving constrained variational problem, the center frequency and bandwidth of each IMF component is constantly updated with the iterative process[11]. For any bearing vibration signal $y(t)$, constrained variational problem can be constructed as follow:

$$\begin{cases} \min_{\{y_k\}, \{w_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times y_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^K y_k(t) = y(t) \end{cases} \quad (4)$$

Among them, $\{y_k\} = \{y_1, y_2, y_3, \dots, y_K\}$ and $\{w_k\} = \{w_1, w_2, w_3, \dots, w_K\}$ are the IMF components and a collection of center frequency respectively. Using Lagrange factor $\lambda(t)$ and the quadratic penalty factor α to transform the above constrained optimization problems into unconstrained optimization problems, the augmented Lagrange form can be expressed as follow:

$$\begin{aligned} \zeta(\{y_k\}, \{w_k\}, \lambda) &:= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times y_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ &+ \left\| y(t) - \sum_k y_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_k y_k(t) \right\rangle \end{aligned} \quad (5)$$

Formula (5) can be solved by alternate direction multiplier method (ADMM), which solve the saddle point by alternative update $y_k^{n+1}, w_k^{n+1}, \lambda(t)^{n+1}$. The solution of variational problem $y_k^{n+1}(t)$ in frequency domain can be expressed as follow:

$$\hat{Y}_k^{n+1}(w) = \frac{\hat{Y}(w) - \sum_{i \neq k} \hat{Y}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \quad (6)$$

Where $\hat{Y}(w)$, $\hat{Y}_i(w)$, and $\hat{\lambda}(w)$ represent the Fourier transform of $y(t)$, $y_i(t)$ and $\lambda(t)$ respectively, in the formula (6) contains a set of wiener filter. Each time domain signal of IMF components can be obtained through to calculate Fourier inverse transformation of the signal after filtering for real component. Each the center frequency of updated IMF component can be obtained from following formula:

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{Y}_k(w)|^2 dw}{\int_0^\infty |\hat{Y}_k(w)|^2 dw} \quad (7)$$

To construct the simulation signal as follows:

$$x(t) = x_1(t) + x_2(t) + x_3(t) \quad (8)$$

$x_1(t)$ is the cycle impact function, each period impact composition is $0.5 \sin(4000\pi t) \times e^{-(t-t_0)^2/\tau}$, a total of five cycles. $x_2(t) = \sin(1200\pi t)$, $x_3(t) = \sin(400\pi t)$. $\tau = 6 \times 10^{-7}$ s, $t_0 = 0.005$ s, $0 \leq t \leq 0.05$. The sampling frequency is 30 kHz. The simulation signal and its components as shown in figure 5:

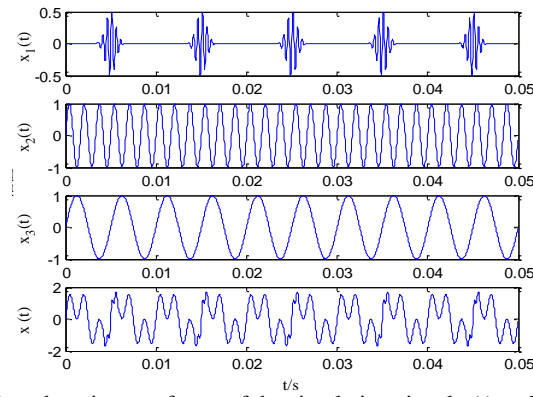


Figure.5 The time domain waveforms of the simulation signal $x(t)$ and its components

To decompose the simulation signal by VMD method and the time domain waveform is shown in figure 6. VMD successfully separate the individual component of simulation signal, each pattern signal is very close to simulation signal.

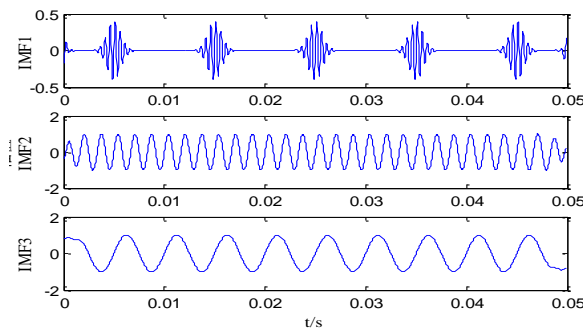


Figure. 6 Decomposition results of VMD of the simulation signal

4. BEARING FAULT DIAGNOSIS METHOD BASED ON STOCHASTIC RESONANCE AND VMD

The stochastic resonance method is used to denoise the bearing vibration signal for extracted from acceleration sensors. Then use VMD to decompose the de-noised signal and get a number of IMF components. Finally, the fault characteristic frequency can be found in each IMF component spectrum to determine the fault type of the bearing. The fault diagnosis flow chart is shown in figure 7.

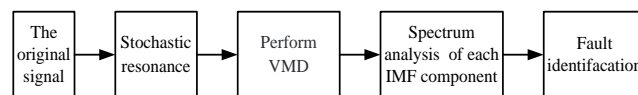


Figure. 7 Fault diagnosis process based on stochastic resonance and VMD

5. THE TEST SIGNAL ANALYSIS

For the mechanical equipment of continuous operation, bearing is under a large load in the continue working process and it easy to appear all kinds of faults. The early fault feature signals are weak and tend to be submerged by the noise in the working environment so that it is difficult to find fault. But it can be used stochastic resonance method to extract characteristic signal.

The bearing vibration signal data from rotating machinery diagnostic laboratories, the type of deep groove ball bearing is ER-10k, The roller diameter is 7.9 mm, Pitch circle diameter of roller is 33.5 mm, number of roller is 8, the contact Angle is 0, experimental setup consists of a 1/3 horsepower 3 phase motor, a coupling, a tachometer and the corresponding electrical control device. Test rig structure as shown in figure 8, The motor speed is 3300r/min. acceleration sensor respectively installed on the corresponding point of the bearing by vertical and radial, vibration signal can be collected from eight channel ZonicBook/618E

data recorder, Sampling frequency is 2560Hz. According to the parameters of the bearing, the fault characteristic frequency of Inner-race and rolling element be calculated as shown in table 1.



Figure.8 Experimental setup of bearing test rig

Table 1: The Characteristic frequency of rolling bearing components (Hz)

| Shaft rotation frequency | inner-race fault frequency | rolling element fault frequency |
|--------------------------|----------------------------|---------------------------------|
| 55 | 272.25 | 109.56 |

5.1. Inner-race fault signal analysis

The 5120 points were analyzed under the inner-race fault. The time domain waveform of fault signal is shown in figure 9. It is difficult to identify the impact composition of the bearing fault signal From the time domain waveform, because the signal be submerged by strong background noise.

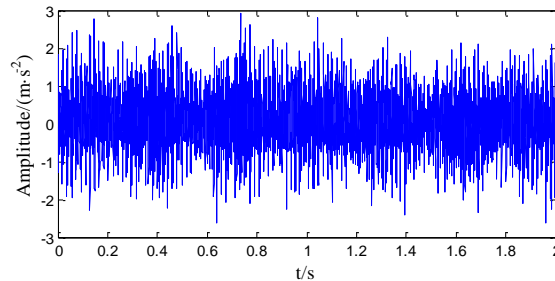


Figure. 9 The time domain waveform from inner-race fault

Figure 10 is time domain and frequency domain waveform that decomposed by VMD of inner-race fault signal. From the frequency domain of figure 10 can clearly see the spectrum lines of shaft rotation frequency and 3 times fault characteristic frequency.

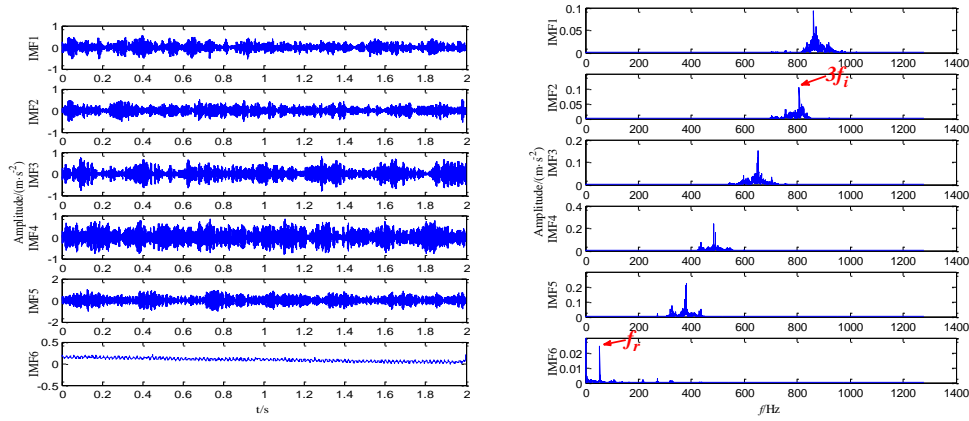


Figure. 10 Decomposition results of VMD

The output signal time domain waveform of stochastic resonance is shown in figure 11. The decomposition results of VMD is shown in figure 12. From the frequency domain of figure 12 it can clearly see the spectrum lines of fault characteristic frequency and its 3 times frequency. The fault characteristic spectral lines in figure 12 is more clearly than figure 10.

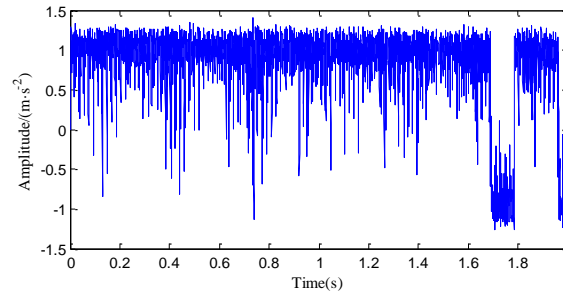


Figure. 11 The output signal time domain waveform of stochastic resonance

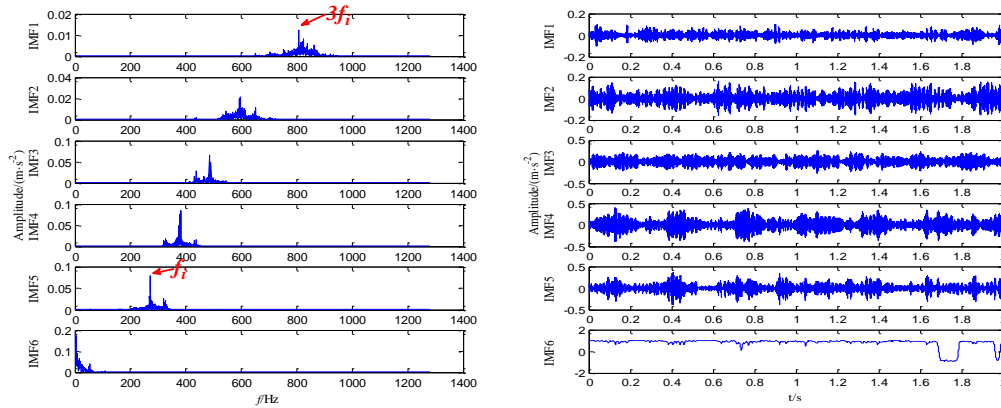


Figure. 12 The output signal of stochastic resonance decomposition results of VMD

5.2. The rolling element fault signal analysis

The 5120 points were analyzed under rolling element fault fault. The time domain waveform of fault signal is shown in figure 13.

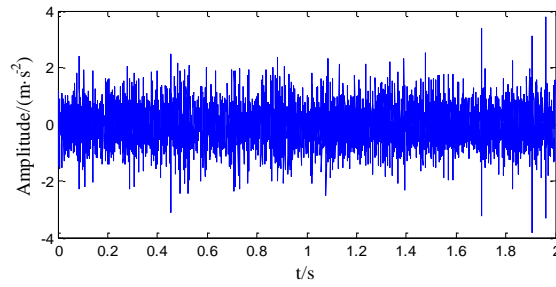


Figure. 13 The time domain waveform from rolling element fault

Figure 14 is time domain and frequency domain waveform that decomposed by VMD of rolling element fault signal. From the frequency domain of figure 14 can see the spectrum lines of fault characteristic frequency and its 4 times frequency

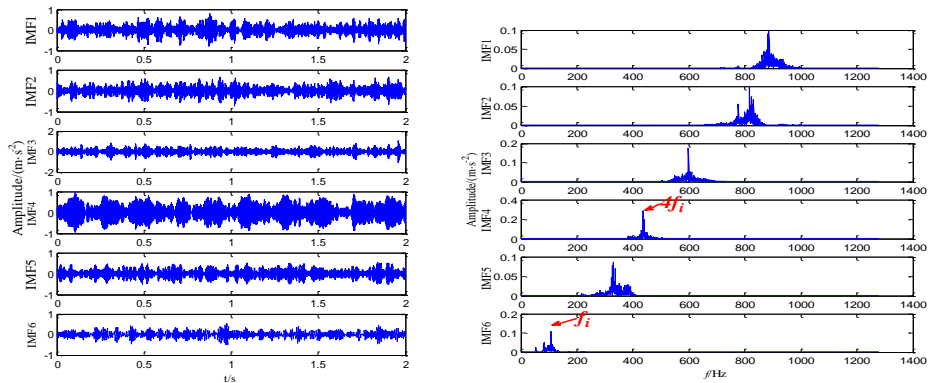


Figure. 14 Decomposition results of VMD

Figure 15 is the output signal time domain waveform of stochastic resonance. Figure 16 is time domain and frequency domain waveform of the output signal of stochastic resonance decomposed by VMD. From the frequency domain of figure 16 can clearly see the spectrum lines of fault characteristic frequency and its 4 times frequency. The fault characteristic spectral lines in figure 16 is more clearly than figure 14. So the quality of signal decomposition by VMD can be improved by stochastic resonance.

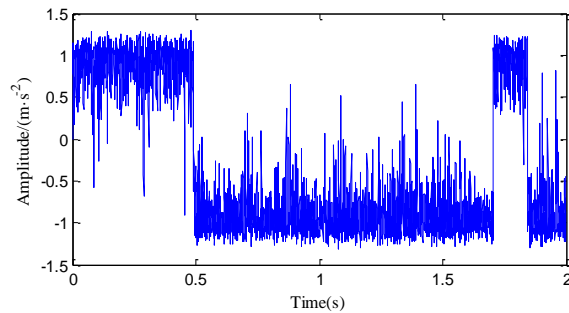


Figure. 15 The output signal time domain waveform of stochastic resonance

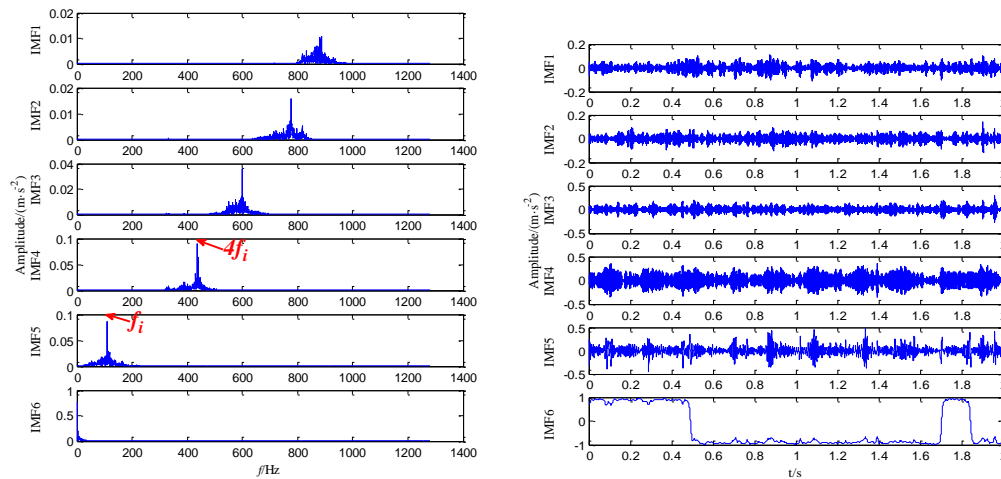


Figure. 16 The output signal of stochastic resonance decomposition results of VMD

6. CONCLUSION

The working state identification of bearing has been the core issue of rotating machinery fault diagnosis, however the decomposition quality and accuracy of VMD is seriously affected by the strong background noise. Aiming at the problem that a fault diagnosis method based on stochastic resonance and VMD is proposed in this paper. The experiment results indicate that the method has a good prospect in detection weak signal of the bearing early fault.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (51565046), Natural Science Foundation of Inner Mongolia Autonomous Region, China(2015MS0512), Colleges and Universities Research Project of Inner Mongolia Autonomous Region, China (NJZY146).

REFERENCES

- [1] I. El-Thalji, E. Jantunen (2015). A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*. Volumes 60-61, 252-272.
- [2] X. Y. Zhang, Y.T. Liang, J. Z. Zhou, et al (2015).. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, Volumes 69, 164-179.
- [3] R. Benzi, A. Sutera, A. Vulpiani, (1981) The mechanism of stochastic resonance. *J. Phys. A: Math. Gen.* Volume 14, Issue 11, 453-457.
- [4] J.M. Li, Y.G. Zhang, P. Xie(2016). A new adaptive cascaded stochastic resonance method for impact features extraction in gear fault diagnosis. *Measurement*. Volume 91, 499-508.
- [5] G.Y. Li, J.M. Li, S.B. Wang, et al (2016). Quantitative evaluation on the performance and feature enhancement of stochastic resonance for bearing fault diagnosis. *Mech.Syst.Signal Process.* Volume 81, 108-125.
- [6] Z.H. Lai, Y. G. Leng (2016). Weak-signal detection based on the stochastic resonance of bistable Duffing oscillator and its application in incipient fault diagnosis. *Mech.Syst.Signal Process.* Volume 81, 560-74.
- [7] Dragomiretskiy K, Zosso D (2014). Variational mode decomposition[J].*IEEE Tran on Signal Processing*. Volume 62, Issue 3, 531-544.
- [8] X.L. An, J.J. Yang (2015). Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy. *Transactions of the Institute of Measurement and Control*. 531-544.
- [9] Z.L. Lv, B.P Tang, Y. Zhou, et al (2016). A Novel Method for Mechanical Fault Diagnosis Based on Variational Mode Decomposition and Multikernel Support Vector Machine. *Shock and Vibration*. Volume 2016, Article ID 3196465, 11 pages.
- [10] J Zhu, C Wang, Z Y Hu, et al (2015). Adaptive variational mode decomposition based on artificial fish swarm algorithm for fault diagnosis of rolling bearings. *Proc IMechE, Part C : J Mechanical Engineering Science*. Volume 0, Issue 0, 1-20.
- [11] C Aneesh, K Sachin, P M Hisham, et al (2015). Performance comparison of Variational Mode Decomposition over Empirical Wavelet Transform for the classification of power quality disturbances using Support Vector Machine. *Procedia Computer Science*. Volumes 46 372-380

Session 5

Maintenance Engineering / Risk, Safety Assessment and Management

Session Chair Dr. Guanming Dong

How Can SMES Adopt a New Method to Advanced Maintenance Strategies? A Case Study Approach *(D. Baglee, U. Gorostegui, E. Jantunen, P.Sharma, J. Campos)*

The Role of Cultural Development When Improving Maintenance Practice in the Automotive Supply Chain *(D. Dixon, K. Robson, A. Wheatley, D. Baglee)*

The role of ISO 55000 Standard in Asset Integrity
(I. Ihemegbulem, D. Baglee, A. Wheatley, S. Mattonen-Arola)

How Can SMEs Adopt a New Method to Advanced Maintenance Strategies? A Case Study Approach

D. Baglee¹, U. Gorostegui², E. Jantunen³, P. Sharma⁴, J. Campos⁵

³VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Finland

ABSTRACT

Maintenance is crucial to manufacturing operations. In many organisations, the production equipment represents the majority of invested capital, and deterioration of these facilities and equipment increases production costs, reduces product quality. Over recent years the importance of maintenance, and therefore maintenance management, within manufacturing organisations has grown. The maintenance function has become an increasingly important and complex activity, particularly as automation increases. The opportunity exists for many organisations to benefit substantially through improvements to their competitiveness and profitability by adopting a new approach to maintenance management. Several tools and technologies including Condition Based Maintenance (CBM), Reliability Centred Maintenance (RCM) and more recently e-maintenance have developed under the heading of Advanced Maintenance Strategies. However, the adoption of advanced maintenance strategies and their potential benefits are usually demonstrated in large organisations. Unfortunately, the majority of organisations are constrained by the lack of knowledge and understanding on the requirements, which need to be in place before adopting an advanced maintenance strategy. These are usually classified as Small and Medium Sized Enterprises (SMEs).

The research strategy is based on ‘empirical iterations’ using survey secondary data, experts’ interviews information and multiple case studies. The results show that there is a set of recommendations, which strongly influence the implementation of an Advanced Maintenance Strategy (AMS) with a Small to Medium Enterprise (SME). Organisations require a structured and integrative approach in order to take advantage of a new approach to maintenance management. This paper will propose recommendations for integrating an AMS into the organisation and provide evidence of a successful implementation.

Keywords: Condition Based Maintenance (CBM), Strategy Development

Corresponding author: Dr David Baglee (david.baglkee@sunderland.ac.uk)

1. INTRODUCTION

Today, Europe's manufacturing industry directly contributes around 22 per cent of EU GDP [9]. As a result, many analysts believe that if Europe is to become the world's most competitive knowledge-based economy by 2020, the presence of a strong and competitive manufacturing sector is indispensable. In addition, the global economic environment requires manufacturers to adapt to a volatile climate and find new ways to maintain stability and drive competitive advantage. One such way is to develop and implement a maintenance strategy, which is appropriate, cost effective, and energy efficient. Therefore, benefits from a sustainable resource management strategy would see a reduction in manufacturing costs and a reduction in energy usage and wastage, thus allowing manufacturers to make better use of resources to achieve greater cost savings and to prepare for an increasingly regulated environmental landscape. Sadly, relatively few companies either see, or take, this opportunity. Why?

Part of the answer lies in ignorance – especially in the small company sector. Many SMEs simply are not aware of how to do maintenance better – or if they are, they think that it will cost them too much to do so. This perception flies in the face of the facts. There are many, many examples of companies (including small

companies) reaping generous rewards from doing maintenance better – reduced costs, improved availability and reliability, better quality, and more profit. Return on Investment can, and should, be a major factor in making changes to maintenance practices, but this is not apparent to most companies who seem to be afraid of making the investment necessary to reap the rewards.

Over recent years, the importance of maintenance, and therefore maintenance management within European manufacturing organisations has grown. In addition, companies are under pressure to respond to rising energy costs and the need to protect the environment. Growth in energy consumption has a direct impact on the deterioration of the environment and on climate change. Air quality is a major environmental concern for the EU. The Commission is currently elaborating the EU Clean Air Programme (CAFE), where the harmful effects of ozone and especially particulates are revealed for human health and ecosystems.

European manufacturers have introduced a variety of innovative technologies, new business processes and enlightened management techniques to encourage greater efficiency in the industrial use of energy. However, equipment maintenance has been overlooked and falls short with regard to the development of innovative and new technologies to monitor energy and support the development of maintenance strategies. The most important barrier to increased energy efficiency is a lack of information (on costs and availability of new technology; lack of information on costs of own energy consumption; lack of training of technicians on proper maintenance and the lack of information and training on the latest technologies and their economic and financial impact on the rate of return from investment). Investments in cost-effective energy efficiency improvement will almost always have a positive impact on manufacturing costs, energy consumption, energy costs, skill development and the environment by reducing the amount of energy consumed and wasted through poor and inadequate bad maintenance practices. Table 1 provides an overview of costs associated with lost production in one year attributed to unplanned downtime.

Table 1: examples of costs of lost production or unplanned downtime [1].

| Manufacturing Sector or Activity | Average Cost of One Day's Lost Production or Unplanned Downtime |
|-----------------------------------------|------------------------------------------------------------------------|
| Automated Production Line | 100,000 € |
| Automotive Engine Production Line | 288,000 € |
| Chemical Production Facility | 100,000 € |
| Paper or Pulp Plant | 200,000 € |
| Power Plant | 100,000 € |
| Cargo Ship | 10,000 € |

Research [1,2,3] have stated that a methodology for maintenance strategy development needs to incorporate an appropriate mix of maintenance practices and technology relevant to manufacturing taking into account the development of energy efficient maintenance practices. This will allow the company to focus on developing a sustainable methodology utilising modern maintenance strategies, which are multi-dimensional, such as Total Productive Maintenance, Reliability Centred Maintenance and Condition Based Maintenance to allow management to develop and implement a modern maintenance methodology which suit their often perceived constraints such as a lack of financial support, a lack of skilled operators and a lack of time to develop and implement alternative strategies.

In addition, a structured implementation process needs to be mapped to the corporate strategy to suit the company requirements. This will require the development and validation of pilot models enabling the implementation of the models and demonstrate the benefits resulting from the developed appropriate maintenance strategies. The sustainable models will incorporate managerial and technological advances for identifying, developing and implementing a complete asset management system based upon a Return on Investment (ROI).

Maintenance of plant assets has acquired an importance that it never had. In the current economic downturn worldwide, plant asset maintenance has become equally important as the production itself. Money saved with optimum maintenance is money earned. Industries no more see expenditure on maintenance as wasteful. Maintenance costs have now become investments that yield good results. There is a cost

associated with diagnostic actions. An increased inspection frequency results in higher incurred cost for the industry. Therefore, optimization of diagnostics schedule is an issue that needs careful attention. Deciding optimal condition monitoring intervals is another issue needed to be addressed. The condition-based maintenance of large assets of an industry should be carried out with seamless integration of data and information with other activities (such as production/inventory planning) that are influenced by maintenance actions so that optimum utilization of plant assets is effected with minimum overall cost. Intelligent Maintenance Management Systems (IMMS) have emerged as the solution to address these complex requirements.

It helps the maintenance managers to achieve highest levels of plant asset performance by shifting from a reactive “fix when failed” approach to a more proactive “predict and prevent” approach. This requires them to synchronize maintenance activities with production schedules through the use of information technology i.e. wireless, web, etc. [6]. This requires heavy reliance on acquiring and transforming machine condition data to useful information for improved productivity and asset utilization [8]. The use of e-technologies increases the possibilities of utilizing data of multiple origin and type, processing larger volumes of data and to make more advanced reasoning and decision-making, and implementing cooperative (or collaborative) activities (Marquez, 2007). This section will make an effort to highlight the current state of Maintenance Management Systems, e-maintenance and its advantages and describes the need to have ontology as knowledge repository.

2. CURRENT MAINTENANCE MANAGEMENT SYSTEMS

Most manufacturing industries today have realised the importance of going beyond the conventional maintenance management techniques to computerised maintenance management. The increase in processing powers of computers, coupled with a sharp decrease in the cost of acquiring them has led to a surge in use of computers in the field of maintenance management as well. Computer Maintenance Management Systems are being used with limited success in many industries today. Most of these management systems are limited in their capabilities where they help generate work orders and job sheets, assign maintenance personnel to their tasks, manage inventory to a certain extent and produce maintenance schedules based on set rules. But there has been a tremendous increase in the complexity of the manufacturing operations, which in turn demands a lot more from the investments being made in the maintenance management systems. These systems do not have the capability of knowledge storage and retrieval, prognosis or diagnosis. Some Maintenance Management Systems have moved ahead in this sense by having certain ‘plug and play’ Condition based Maintenance modules being supplied by different vendors. CBM modules are available which enable a CMMS to use condition monitoring alarm levels to trigger maintenance activities.

Incoming condition-based data for assets is compared to predefined thresholds and when the threshold is exceeded an alarm is raised to highlight the event. [5] (2006) have highlighted another improvement in the current Maintenance Management Systems. Some of these have criticality assessment capability in them. This capability helps in assessing the parts of the manufacturing system that are most critical to the operation. This is arrived at by evaluating the past data and collective knowledge of the operators. This in turn leads to formulation of maintenance strategy with these critical assets as the focus of attention. These improvements have led to different set of problems arising. [4] (2005) observed that with the numerous asset management systems offered by different vendors, the process of integration can be problematic as many systems have their unique data exchange interfaces. This is a common problem being faced by many plants and factories. It often so happens that the customisation of the needs of a plant requires more than one system and more often than not, from different vendor. Purchasing systems from a single vendor leads towards system compatibility, however suppliers may not provide a total asset management solution, and the reliance on one vendor can prove risky. This leads to a conflict due to differences in data exchange interface. Custom bridges are available to integrate various systems being supplied by various vendors but that means additional expense. Also, these custom bridges need updating as and when the vendor brings out a newer version of his asset management system. Another option is to use an industry-standard bridge,

which allows businesses to mix different systems with reduced integration costs. However, there may be performance loss compared to a custom solution and vendors must be willing to support the standard.

A set of ISO standards deals precisely with addressing the problems of integration of modules of different vendors. ISO 16100 aims to provide a method to represent the capability of manufacturing application software relative to its role throughout the life cycle of a manufacturing application, independent of a particular system architecture or implementation platform. This can lead to reduced production and information management costs to users and vendors/suppliers of manufacturing applications. The reasons for developing ISO 16100, as given in www.iso.org are as given below.

- a growing base of vendor-specific solutions;
- user difficulties in applying standards;
- the need to move to modular sets of system integration tools;
- the recognition that application software and the expertise to apply that software are assets of the enterprise.

ISO 15926 is an International Standard for the representation of process industries facility life-cycle information. This representation is specified by a generic, conceptual data model that is suitable as the basis for implementation in a shared database or data warehouse. The International Standard (ISO 15745) defines an application integration framework - a set of elements and rules for describing integration models and application interoperability profiles; e.g. generic elements and rules for describing integration models and application interoperability profiles, their component profiles - process profiles, information exchange profiles, and resource profiles. ISO 18435 provides a framework for harmonized use of industry and international standards in order to integrate control, diagnostics, prognostics, capability assessment, and maintenance applications. By using an ISO 15745 application integration modelling approach, key interoperability interfaces can be identified and concisely documented in terms of profiles. ISO 18435 also provides the elements and the rules to describe the integration requirements of an automation application. The elements include the key aspects when integrating an automation application with other applications and the relationships of these key aspects.

3. CONDITION BASED MAINTENANCE

CBM is a predictive maintenance strategy that focuses on monitoring the critical parameters of an asset to determine its current health level and predict future problems to avoid them beforehand by scheduling maintenance at the optimal moment. It is worth noting that implementing a CBM system is a continuous process more than a single step plan.

To be able to implement this maintenance strategy, it is necessary to do an initial study about the asset that is going to be monitored to decide whether it is cost effective or not. However, it is possible that not every asset is suitable for a CBM system. Usually, CBM is used for safety critical components where an unexpected failure could cause a risky situation or could result in costly losses.

After the initial cost analysis, an equipment audit needs to be carried out, to choose which parameters generate the most meaningful information. Once this is analysed and the parameters decided, a suitable set of transducers is chosen to monitor said parameters. The gathered data are consequently manipulated to extract the significant descriptors that help diagnose the current health level of the asset. These descriptors are then compared with the historical data and, with the assistance of mathematical and physical models of the system, it is possible to predict how the asset is going to behave in the future. The models for prognostics can be classified into two main groups: data-driven prognostics and model based prognostics. The data-driven prognostics are based on pattern recognition from previously collected data, and while it could be appropriate for very complex systems, obtaining initial run-to-fail data could be time costly. On the other hand, the model based ones create a physical model of the system based on mathematics and

physical interactions. Nonetheless, this approach balances between the coverage and the accuracy of the model. Typically, the designed prognostics system is a hybrid between both approaches.

The key concept in a CBM system is the recommended maintenance action. Once the system predicts the future behaviour of the asset, it should create an action report containing important information to link it with the CMMS.. This action reports could include instructions on disassembling the component, the tools needed or the priority of the repair. This way, if it is known when a component is going to fail, it is possible to schedule the pertinent maintenance actions so that it causes the least downtime possible, increasing the availability of the asset and consequently the Overall Equipment Efficiency (OEE). It could make automatic orders for new components so that by the time the technician gets to the destination, the spare part is already there.

All the process can be done remotely and in real-time if the system is implemented in a cloud service. The users can access the information from any device with an internet connection, making the system very versatile. Microsoft Azure, Google Cloud or Oracle offer this type of service.

4. BACKGROUND TO CASE STUDY

This case study provides an overview of the development and implementation of a maintenance concept within company A. who are a subsea cable manufacturer in Europe. Senior management at Company A are aware that maintenance contributes significantly to the total cost of production, and that an effective maintenance system is an efficient way of enhancing a company's capability to handle production losses due to unplanned downtime. However, the introduction of an improved maintenance system would require changes, and these changes, especially changes to the corporate culture, always meet resistance and generate conflict.

Therefore, the management at Company A, required a maintenance concept tailored to their particular organisation, one which would help to identify the barriers and obstacles such as lack of resources, lack of skills and cultural change, which may impede the integration of a new maintenance concept. However, in order to be effective the concept would need to have a supporting structure/model around which a detailed maintenance plan could be developed. To implement the framework or model successfully, certain change factors, identified through research undertaken, at the University of Sunderland, would need to be addressed. The factors are:

- create support for change from top management
- formulate a clear policy or strategy form improved maintenance activities
- explain the meaning and extent of the new initiative to everybody involved with maintenance
- provide a strong organisational responsibility for supporting the implementation
- promote the new initiatives through group activity

The framework, developed in partnership with the management at Company A was designed by identifying which elements were essential to allow the asset manager to develop a maintenance strategy which: (1) Records data relevant to the cost of failure and the cause and effect of the failure. (2) Records the estimated cost of the current maintenance system (3) provides a detailed list of tasks ranging from basic inspection to fault diagnosis allowing the user to determine which tasks to select based upon equipment needs and the resources available. (4) Provides an estimate of the cost to implement the sub elements selected.(5) Provides a ROI analysis to determine if the new approach is feasible.

The model, Advanced Integrated Maintenance Management System (AIMMS) is structured in a way that allows the management to follow a 'path' or 'road map' to the implementation of an appropriate maintenance initiative. The model, which can be viewed in figure 1 offers Company A management a fundamental paradigm shift from traditional activities, to the adoption of a tailored system that incorporates acceptable and effective elements of modern maintenance practices such as Total productive Maintenance

(TPM) and Reliability Centred Maintenance (RCM). It should be noted that that this model is a 'management tool' and is designed to complement a Microsoft Access database.

5. MODEL DEVELOPMENT BASED UPON THE KEY ELEMENTS IDENTIFIED AT CASE STUDY COMPANY

The model contains a decision support module aimed to increase the effectiveness of the maintenance function. However, in order to utilise this module and develop an improved maintenance strategy the maintenance management needs a powerful tool, information. Information is readily available, and raw data is collected, analysed and processed. After this process it becomes significant for maintenance planning and decision-making. The collection and analysis of data helps the management formulate a strategy, which is based upon real and accurate data. The model allows the data to be stored and used within different scenarios, therefore providing an alternative approach, if required.

A new concept is not always clear or understood, and hence it was thought that the maintenance engineers would not utilise the system and stick with their own way of doing things. Therefore, in order to be successful the maintenance department would need to be consulted about the different approach to maintenance and how the model would work in conjunction with the Database already developed. Meetings were held with maintenance engineers and management and the problems and possible solutions were discussed. The use and requirements of the model were discussed with management, comments were used to re-design and improve the model.

6. MODEL EVALUATION

Certain difficulties were identified, the management stressed that a precondition for change would have to be an immediate increase in OEE and an increase in operator maintenance. The increase in OEE was evident within 4 weeks, although the increase in operator maintenance was not. The argument from the maintenance department was that the model allowed the operator to input a suggestion for maintenance but did not take into account the fact that operator maintenance initially requires extra support from maintenance staff, providing help with the extra workload of the initial phase. Therefore, without a push for change on behalf of the maintenance engineers, success if any would be limited, and if the maintenance department were busy the operators would have to wait or seek alternative help.

The model had the capability to be linked to the Database produced by Assynt and the University. The database includes PowerPoint slides to develop single point instructions and training exercises, the model allows the staff to develop a maintenance strategy, and if this strategy is successful the new approach could be translated into PowerPoint slides or other training devices, thereby increasing the capacity for training.

The analysis of the equipment data has identified that the first stage evaluation was:

- useful and helped to develop an alternative maintenance strategy and is reasonably easy to use and understand
- flexible enough to be used on different equipment from within the different bottling lines
- able to identify alternative maintenance approaches
- provides a progressive approach to implementation
- can be used to educate and train the workforce
- provides an identification of the cost of maintenance

The model, after the first stage evaluation would appear to be capable achieving its intended aims.

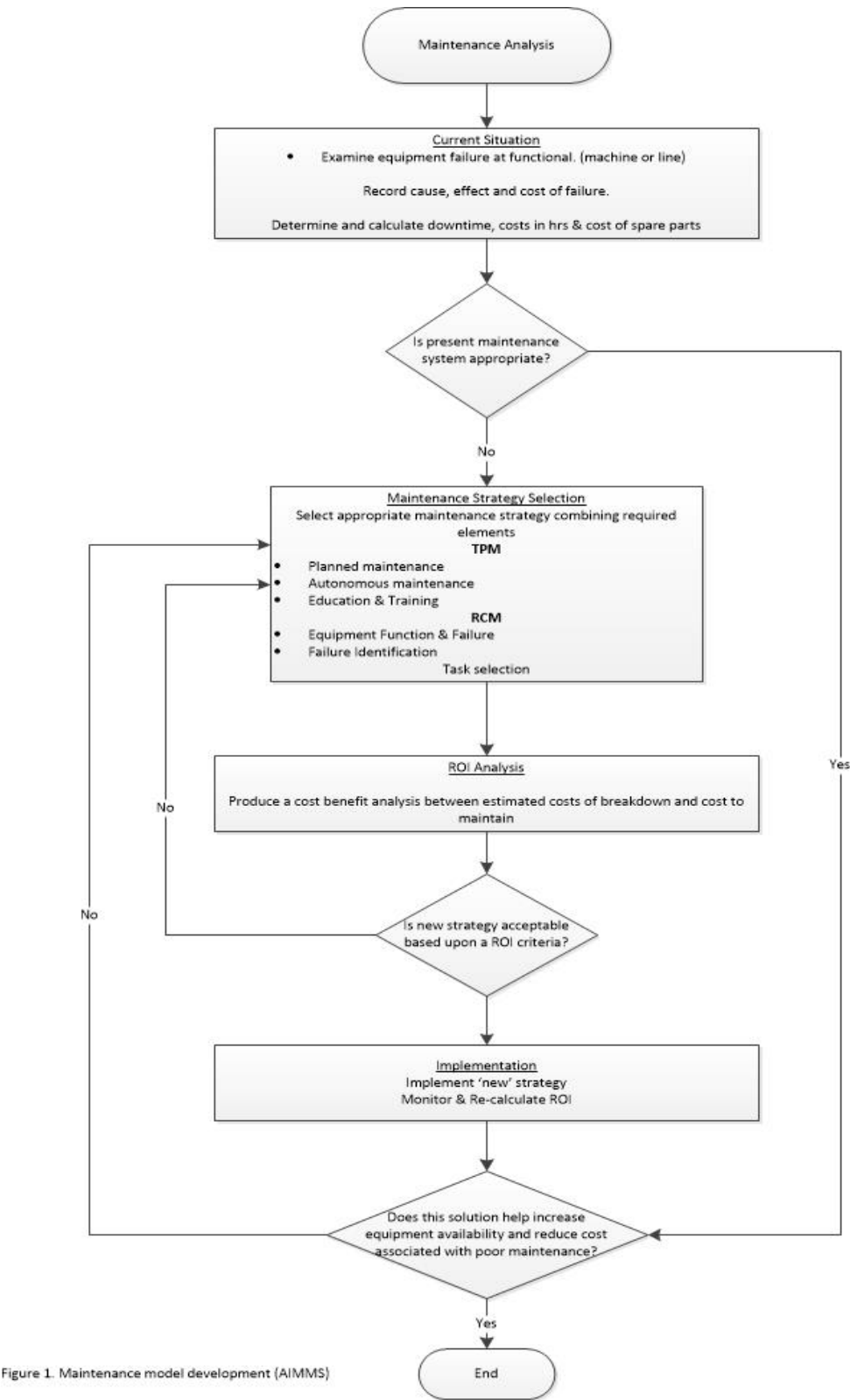


Figure 1 Maintenance model development (AIMMS)

7. DISCUSSION

In some cases, the equipment is complex, yet it should be too difficult to analyse with a view to generating continuous improvement. The database, developed in partnership with Assynt and the University, and the model developed in partnership with Company A and the University

Operators were sceptical about the possibility of extra work and the maintenance engineers believed they would not have enough say, and a new method to hopefully improve maintenance may indicate that the maintenance and operators were not doing their job properly. However, the model was not used to identify the failures or shortcoming of staff but to aid the management to perform machinery analysis of losses on a detailed level. The model was also developed with the aim of promoting technical changes to the production system, by promoting operator maintenance through condition monitoring. Finally, the model managed to achieve, albeit on a small scale, a balance between an identification of the technical complexity, the identification of the competence of the operator and an increase in the amount of access the operators had to the maintenance engineers.

8. CONCLUSION

As a result, companies' performance and responsiveness will be substantially increased, and both management and workforce knowledge/skills will be consistently and continuously developed. The project will also contribute to the reduction in the release of harmful gases and toxins by developing systems which allow for the analysis of energy consumption i.e. water, electricity, and raw materials. In addition, Maintain-IT will benefit small local industry by developing a "network of excellence" between academia and industry to create a forum for debate to encourage innovative thinking and develop new methods for maintenance and manufacturing.

REFERENCES

- [1] Baglee, D. Knowles, M. (2013) "Maintenance Strategy Development in the UK Food and Drink Industry" – International Journal of Strategic Engineering Asset Management - Vol. 1, No.3. pp 289 – 300
- [2] De Groote, P. (1995) "Maintenance performance analysis: a practical approach", Journal of Quality in Maintenance Engineering, Vol. 1 Iss: 2, pp.4 – 24
- [3] Makarand, S. Kulkarni. (2013) Optimisation of opportunistic maintenance of a multi-component system considering the effect of failures on quality and production schedule: A case study The International Journal of Advanced Manufacturing Technology Volume 69, Issue 5-8 , pp 1743-1756
- [4] Mathew, A.D., Zhang, L., Zhang, S. and Ma, L., (2006), "A review of the MIMOSA OSA-EAI database for condition monitoring systems", In Proceedings World Congress on Engineering Asset Management, Gold Coast, Australia.
- [5] Moore, W. J. and Starr, A. G., (2006), "An Intelligent Maintenance System for Continuous Cost-Based Prioritization of Maintenance Activities", Computers in Industry, Vol. 57, No. 6, pp. 595-606.
- [6] Muller A., Suhner M.C. and Iung B., (2004), "Probabilistic vs. dynamical prognosis process based e-maintenance system", IFAC-INCOM'04, Information Control in Manufacturing, Salvador, Brasil.
- [7] Ramesh, G. (2013) Does it pay to design for reliability and maintainability? Asset Management & Maintenance Journal, Vol. 26, No. 3, pp39-40
- [8] Rezg, N. and Dellagi, S., (2010), "Intelligent Integrated Maintenance Policies for Manufacturing Systems", Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management, Springer.
- [9] Trimble, R. Baglee, D. MacIntyre, J. "A New Model for Maintenance Management in SMEs. Journal of Quality, Reliability and Maintenance. Professional Engineering Publishing Limited, 2004. Vol 5 Pp. 127-130

The Role of Cultural Development When Improving Maintenance Practice in the Automotive Supply Chain

Derek Dixon¹, Kenneth Robson¹, David Baglee¹, Alan Wheatley¹

¹ Faculty of Engineering and Advanced Manufacturing, University of Sunderland, United Kingdom.

ABSTRACT

The automotive supply chain is an essential contributor to the UK economy. Production of commercial vehicles is expected to rise to an unprecedented 2.1 million in 2018. Contributing organisations are expected to be flexible, responsive and cost efficient which is a direct requirement of the Original Equipment Manufacturer (OEM) operating in a lean production context. To operate effectively within the supply chain relies on each department or function in the business working at maximum efficiency. At a granular level, the operation and deployment of the maintenance facility becomes crucial. Yet the strategy for maintenance deployment can often be influenced and even restricted, by managerial and cultural issues.

These issues are unique to this sector and a direct result of the constant demands of the OEM. They include an over reliance on buffer stock to mitigate the risk of maintenance failure, as well as a negative cultural influence on the business. Moreover, a complete absence of communication or technical support for maintenance across the supply chain, leading to isolated and immature practices. Consequently, suppliers are underperforming.

This paper reports on the challenges and contextual factors, which influence the maintenance strategy for automotive manufacturing suppliers. In addition, a solution will be offered to these damaging issues through the development of a maintenance strategy tool. The tool will look to address identified constraints through cultural development and the utilisation of a series of key performance indicators. These performance indicators will form a Maintenance Performance Measurement (MPM) system.

Keywords: Automotive Supply chain, Organisational Culture, Maintenance Strategy Development, Performance Measurement

Corresponding author: derek.dixon@sunderland.ac.uk

1. INTRODUCTION

The automotive industry in the UK is experiencing a period of growth and success. The subsequent contribution to the local, national and international economy is substantial and the importance of the industry as a whole, is increasing. The supply chain is an essential part of this process, with over 2000 registered automotive suppliers in the UK and approximately 78,000 people employed (1).

Lean production methods and Just in Time (JIT) delivery to the customer, ensure membership of the automotive supply chain is challenging. Furthermore, the success of these organisations depend on all internal departments functioning efficiently (2). Consequently, the importance of the maintenance function operating effectively becomes crucial to the success of the organisation. Recent empirical research by the author within the automotive supply chain identified ineffective maintenance practice and regressive working practices. Furthermore, contrary to lean principles, these companies were holding excessive safety stocks in order to mitigate poor maintenance performance (3). These were unexpected findings, given the automotive context. Furthermore, the results from follow-up interviews pointed to organisational culture as a possible factor affecting maintenance strategy within the case study organisations.

This paper presents findings from case study research carried out with automotive supply chain members in the North East of England. In addition, the role of organisational culture is explored including its effect on strategy across the case study sites.

The paper is comprised of a literature review in Section 2, empirical research findings in Section 3, a potential way forward in Section 4 and Section 5 concludes the paper.

2. LITERATURE REVIEW

This section will begin with a review of literature on maintenance strategies. Subsequently performance measurement is discussed and the section will conclude with a reflection on organisational culture.

2.1. Maintenance strategy development

Maintenance is a fundamental component of an effective manufacturing business. Furthermore, the impact a maintenance function can provide towards the efficiency of a production department is substantial (4). The development or selection of any maintenance strategy must be done with a holistic, business wide perspective as recognised by (4–9). In addition to this perspective, (10) proposed a strategy development model which discussed the need to capture the human component of maintenance. This consisted of tacit knowledge, experience as well as motivation. Additionally, this is recognised by (11) and (12) as being a key consideration in effective strategy development and maintenance management. This is confirmed by (12), who discusses the importance of staff engagement, promoting the benefits to any maintenance strategy as well as to the organisation.

2.2. Automotive Industry specific strategies

Industry specific strategies are not unusual within maintenance management. Total Productive Maintenance (TPM) was formulated for use within the automotive industry in the late 20th century (11). TPM offers substantial department and organisational benefits (13), yet successful implementation requires a holistic approach, employee empowerment, education and training. Additionally, senior management engagement is crucial both financially and emotionally (2,5). (2) confirms that there is no simple recipe for its success, as any such programme cannot predict the skill profile or age range of the employed staff. The programme must be tailored to the needs of the site. Conclusively, the required financial investment for successful deployment can cause conflict with the business, where the specific measurement and financial efficiency of each business is at the very heart of automotive supply chain goals (14).

2.3. Maintenance Performance Measurement

Performance indicators are recognised as being a key component within any maintenance strategy, yet standard metrics will not provide the information that may be required by the senior management team. Several authors, including (15–17) recognise that the senior management team must be involved in the selection and refinement of any Performance Indicator (PI) system. Furthermore, a predefined set of indicators will not be satisfactory to an individual business. Additionally (18) recognised the importance of measurement of the human element within a maintenance department, in areas such as motivation and expertise. A review of EN15341 standards for maintenance performance measurement highlights the fact that there are few indicators for the human or indeed, cultural element.

2.4. Organisational Culture

2.4.1. What is culture?

Organisational culture is an intangible phenomenon and is difficult to quantify. (19) describes it as set of values beliefs and norms but clearly it is more than this. (20) yet discusses culture as being kinetic in its nature and constructed from personal interaction. It also encompasses leadership behaviours and everyday routines. Furthermore, (21) expanded on this by identifying levels of culture. These started with basic assumptions, rising to beliefs, values and attitudes and concluding with tangible artefacts. It is these artefacts, which are visible to any member of an organisation and examples are numerous. They include dress code, the manufactured product, meetings, posters, department layout and even the appraisal system. This exploration of cultural artefacts within a business reveal that even at a superficial level, recognisable items and tangible features are important and can have a positive or negative impact on the organisational culture of a department or business.

Summarising, organisational culture is formed over time and influenced by several factors. The influence of leaders within the organisation on the levels of culture including artefacts is not to be underestimated.

2.4.2. How does culture effect business performance?

The explicit characteristics that lead to a positive and successful organisational culture may be debated within literature, yet (22) highlights the importance of certain features linked to success. This includes the extent to which the culture of an organisation is homogenous. Where differing sub cultures between departments can emerge, and may be necessary, an increasing separation can lead to difficulties with cooperation and working practices. (19) discusses the necessity of an organisation allowing the organic growth of individual sub cultures, pointing to the organisational dangers of imposing a business wide culture. Crucially, (19) stresses the importance of departments with differing cultures, ensuring cooperation through inter-department integration and cooperation. A further feature is cultural strength or consistency, (21). (22) states this feature signposts the extent to which an organisational culture has an effect on the behaviour of its employees. (21) indicates some discernible features of an organisation with a strong culture. A summary of these include:

- Goal alignment. This reveals itself as not only agreement on the aims of the organisation, but also how they are going to achieve those aims.
- High employee motivation. This relies on the assumption that employees enjoy working for an organisation, which is distinctive and clear. As well as this assumption, an indicator can come from an engagement and rewards scheme.
- Experience. A strong culture is able to learn from previous organisational events. This may manifest itself in dealing with manufacturing, marketing or general business led issues.

Following this, (21) indicates these features, or enablers will translate into organisational benefits, whereby a negative culture may have an adverse effect. Tellingly, the strength or weakness of a culture can be directly affected by the turnover of the workforce. A static, loyal workforce can maintain an existent culture but a transient workforce with high staff turnover can dilute established cultural practices.

2.4.3. Changing the culture of an organisation

Deploying a cultural change programme or maintenance strategy can be problematic and painful for an organisation (20–22), yet certain steps to achieve this process are noticeably common with maintenance strategy development. (20) identifies an eight-stage process to establish the required foundation to manage cultural change.

The change programme discussed by (20), has synergy between with the content of maintenance strategy development discussed earlier in Section 2. By deploying the key elements of a successful maintenance strategy can help change the working culture of a maintenance department. Table 1 identifies the common elements found in the processes of Maintenance strategy development and cultural change programmes.

Table 1: Commonality of content between maintenance strategies and cultural change enablers.

| Characteristic | Maintenance strategy development | Cultural and change enablers |
|-------------------------------|----------------------------------|------------------------------|
| Senior management involvement | ✓ | ✓ |
| Clear objectives | ✓ | ✓ |
| Plan | ✓ | ✓ |
| Bespoke development | ✓ | ✓ |
| Communication | ✓ | ✓ |
| Involvement of the employee | ✓ | ✓ |
| Formal training | ✓ | ✓ |
| Feedback | ✓ | ✓ |
| Department Integration | ✓ | ✓ |

2.5. Summary

The similarities between the content of maintenance strategy development and enablers in cultural change programmes demonstrate alignment between these two concepts. The robust development of a maintenance strategy appears to provide further additional benefits to the organisation, through the improvement in cultural strength of the department. The benefits to the organisation of an effective maintenance strategy are confirmed by (7,8,10), yet if the cultural enablers are also part of the maintenance strategy development process, then (21) confirms the organisation is more likely to become more successful. Furthermore, in addition to the key elements listed in table 1, complimentary elements of maintenance strategy and cultural development may be considered where duality does not exist. These include a number of qualitative strategies such as training, skills development, staff turnover and a rewards and discipline scheme. Conclusively, although the quantitative nature of a maintenance strategy is essential for day-to-day operational management of a maintenance department, the inclusion of specific qualitative components which are identified as cultural enablers, may provide additional organisational benefits.

3. EMPIRICAL RESEARCH AND FINDINGS

The data collection for this paper was derived from doctoral research being carried out within the automotive supply chain. The data was primarily qualitative and derived from a range of Tier 1 suppliers. Analysis identified a number of key constraints which were problematic to the development and effectiveness of the maintenance function within the business. The findings are shown in table 2.

Table 2: Constraints to maintenance effectiveness Adapted from (3)

| Constraint profile | Cultural Enablers | Supplier 1 | Supplier 2 | Supplier 3 | Supplier 4 |
|---------------------------------------|-------------------|-----------------|------------------------------|-----------------|---------------------------------------|
| Training | ✓ | ✓ | - | ✓ | - |
| Skills | ✓ | ✓ | - | ✓ | - |
| Staff Resources | ✓ | ✓ | - | ✓ | ✓ |
| Equipment and spares | - | ✓ | - | ✓ | - |
| Production Integration | ✓ | ✓ | ✓ | ✓ | - |
| Maintenance shift system | - | ✓ | - | ✓ | - |
| OEM | - | ✓ | ✓ | ✓ | ✓ |
| Supply chain partner | - | ✓ | ✓ | ✓ | ✓ |
| Parent Company | - | ✓ | ✓ | ✓ | - |
| Senior Management | ✓ | ✓ | ✓ | ✓ | - |
| Organisational Culture | ✓ | ✓ | - | ✓ | - |
| KPI's | ✓ | ✓ | - | ✓ | ✓ |
| Budget | - | ✓ | - | ✓ | - |
| | | | | | |
| Maintenance strategy deployed. | | Reactive | Planned/ Preventative | Reactive | Proactive/ Planned/ Outsourced |

Table 2 provides additional information to the constraint profile of each business. If a constraint is also a cultural enabler, this is indicated as it assists in demonstrating possible cultural impact. Finally, the incumbent maintenance strategy deployed within each supplier provides helpful context.

3.1. Key discussion points

3.1.1. Senior Management engagement and budget

Rich data from each supplier revealed varying levels of senior management engagement with maintenance development. The senior managers of Supplier 1 and Supplier 3 recognised the importance of maintenance to the business, yet were more passive in relation to the performance of the maintenance function. This included acceptance of poor performance measurement, overall strategy development and poor perception of maintenance within the business. Supplier 2 and Supplier 4 agreed on the importance of maintenance to the organisation, but were more proactive in their involvement.

What is of interest is the possible link of senior management engagement and budgetary constraints to the maintenance function. Supplier 1 and Supplier 3 revealed that budgetary limitations were a large constraint towards any supplementary development of the maintenance department. Conversely, Supplier 4 and to a lesser extent Supplier 2 did not indicate any problem with the provided budget for the maintenance function. Interviews revealed the senior management team held the maintenance function in relatively high esteem, valuing its contribution to the business.

An unexpected issue which emerged from discussions, is the use of safety stock. Safety stock is deployed within each business as a buffer to accommodate any small production issues, whilst maintaining JIT or synchronous delivery requirements. Yet further discussions with Supplier 1 and 3 revealed safety stock was being used to mitigate the risk of the maintenance plan being ineffective. The level of safety stock within one site, had a stock value of over one million pounds, yet surprisingly the maintenance plan was reactive with a tightly constrained budget.

3.1.2. Department integration and Culture

Supplier 3 provided insight into the apparent disengagement with production, discussing the negative view the department had taken of the maintenance function. Maintenance was perceived to operate with irregular and disorganised working practices compared to the goal driven, systematic nature of production. This view was not unusual and offered with varying degrees from Suppliers 1, 2 and 3. Interestingly, research revealed Supplier 1 and 3 demonstrated distinct differences with production for several artefacts discussed by (20). These included department placement and management, methods of tracking performance and shift pattern. Supplier 4 offered a different perspective, indicating a more collective approach to production, with shared targets. The issue of culture was raised by Supplier 1 and 3 as being an issue, though not during senior management interviews. Further probing during interviews with other staff members revealed a perceived lack of respect, budget and working conditions afforded to the maintenance function.

3.2. Summary

It appeared that the extensive constraints listed for Supplier 1 and Supplier 3, could explain why a reactive maintenance plan was used in these companies. Additionally, the lack of department integration within these companies can be represented by the distinct differences in culture between the two departments. (19,22) both agree that an overly divergent department culture can cause problems to an organisation. Conversely, the more progressive approach used by Supplier 2 and Supplier 4, strengthened the literary claim of the importance of senior management engagement, supported by an appropriate budget, training and resources. Interestingly, Supplier 4 had few constraints. The extensive senior management engagement along with the close working between partner departments and the right resources, appeared to provide the foundation for a more developed maintenance plan. Furthermore, discussions revealed the business utilised several of the cultural enablers listed within table 1 as part of its normal working practice. These included clear lines of communication, staff engagement, a rewards scheme as well as an extensive training programme. Conclusively, the overall organisation was financially successful, demonstrated by a significant increase in turnover over the previous four years.

The alignment between qualitative constraints identified from ‘rich data’ as well as the cultural enablers detailed within table 1, offered a novel way forward. The departmental and organisational benefits of a strong pervasive culture, which was not entirely homogenous, could alleviate some of the identified constraints and improve the ability of the maintenance function to perform.

4. A WAY FORWARD

Establishing relationships between cultural dimensions from this research within the context of maintenance and the automotive supply chain appears to be important. It provides a possible method of resolving long-standing maintenance performance issues. Table 3 shows these dimensions embedded within strategy tool resulting in a new qualitative and quantitative conceptual model.

Table 3: A Maintenance strategy development tool for the Automotive supply chain

| Constraint | Qualitative features | Quantitative features | |
|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|----------------------|
| | Indicative characteristic (20,21,23) | KPI available (EN15341) | |
| Senior management engagement | <ul style="list-style-type: none"> Identify key project manager to facilitate development Establish reporting mechanism Establish aims and goals. Promote maintenance development to key staff and managers Maintain visible support for development Initiate maintenance planning cycle | | Cultural development |
| Skills and Training | <ul style="list-style-type: none"> Establish apprenticeship programme. Identify training requirements through appraisal system. | | |
| Staff resources | <ul style="list-style-type: none"> Employee engagement Suggestion and reward scheme Consultation and feedback mechanism Length of service and loyalty reward scheme | | |
| Production integration | Establish symbiotic artefacts including: <ul style="list-style-type: none"> Goals, targets, problem solving Communication of goals and achievement through emails, posters Meeting attendance with production facility. Embedded placement of maintenance technicians within production | | |
| Equipment and spares | <ul style="list-style-type: none"> Ensure robust resilience planning and critical spares analysis. | | |
| Budget | <ul style="list-style-type: none"> Identify key resources and justification for plan deployment Link budget requirements to business goals | | |
| Maintenance shift system | <ul style="list-style-type: none"> Assimilate with production shift pattern Integrate with maintenance plan and budget requirements | | |
| Buffer stock | <ul style="list-style-type: none"> Integrate monitoring process of buffer and align to maintenance performance | | |
| | Maintenance strategy development | Maintenance monitoring | |

Table 3 incorporates the issues noted within this research. It provides a conceptual model which can support the performance and advancement of the maintenance strategy within a particular business. Certain

constraints such as ‘OEM’ and ‘parent company’ were omitted, as they were outside the sphere of influence. The model would not be used to develop a maintenance plan, but foster a method of guiding a business towards recognising how it may monitor, develop and enrich its own maintenance strategy.

Senior management engagement aside, each constraint has a quantitative and qualitative element which can be monitored and improved. Qualitative elements have been derived from the rich data and literature. The tool can be used as follows. If the qualitative column is navigated vertically this suggests that the maintenance strategy can be improved through the development of the infrastructure of the maintenance department. Furthermore, vertical travel through the quantitative column will result in an improvement in monitoring of the strategy. Each area of interaction for an inhibitor requires, where possible, qualitative characteristics as well as quantitative monitoring to maximise the ability of the tool to be understood, then used effectively

The crucial nature of the ‘cultural element’ is often cited within maintenance strategy development (10,12,18). A reflection of literature has demonstrated that incorporating this within a strategy tool has proved elusive. Figure 1 demonstrates elements, which collectively, can lead to cultural strengthening within the business. The horizontal travel of key cultural aspects Senior Management, Skills and Training, Staff Resources and Production Integration will look to improve the cultural position of the department. As confirmed by (22) a shift in cultural strength as well as improving, the homogenous nature of both the maintenance and production departments will lead to organisational benefits. This may be confirmed by the performance of Supplier 4.

Due to the volume of indicators available, a colour code is utilised to indicate the extent of indicators which may be available with the EN15341 standards. The colour coding is simplistic in nature:

- Green indicating extensive indicators available to measure performance of that element
- Amber indicating limited metrics available. Moreover, amber is used to indicate the limited application of available indicators.
- Red provides an obvious demonstration that no indicators are suitable or available within the standard.

Finally, the model will include a PI to monitor buffer stock levels. This has been developed as part of this research and will provide an ongoing measure of maintenance improvement. It is anticipated this improvement will be welcomed in businesses where buffer stock is excessive.

5. CONCLUSIONS AND FURTHER WORK

The findings of this research have demonstrated that continued underperformance of the maintenance department within the automotive supply chain might be linked to a lack of cultural development. The acknowledgement of key cultural enablers, such as senior management engagement, training, employee involvement and staff development have duality with recognised maintenance strategy elements. This duality imposes greater significance on these factors within maintenance development and further detailed investigation within this research is crucial. Additionally, acknowledging the contributing factors that will lead to department integration is important as it improves the ability of maintenance perform successfully. Conclusively, cultural development of a department and organisation appears to have a significant effect on the success of the business. The proposed conceptual tool acknowledges and incorporates these factors, providing aligned performance indicators with qualitative characteristics for each constraint. Further research is required to refine the model, thereby identifying and proposing a suitable selection of characteristics and indicators for use within the supply chain. These include addressing buffer stock levels as an indicator of maintenance performance.

REFERENCES

- [1] Growing-the-UK-auto-supply-chain-March-2015.pdf [Internet]. [cited 2016 Feb 19]. Available from: <http://www.automotivecouncil.co.uk/wp-content/uploads/2015/03/Growing-the-UK-auto-supply-chain-March-2015.pdf>
- [2] Wireman T. Total Productive Maintenance. 2Rev Ed edition. New York: Industrial Press, Inc; 2004. 224 p.
- [3] Dixon D, Robson K, Harrison A, Baglee D. Improving automotive supply chain performance through maintenance strategy development. In: EuroMaintenance 2016. Athens, Greece; 2016.
- [4] Kumar U, Galar D, Parida A, Stenström C, Berges L. Maintenance performance metrics: a state- of- the- art review. Kumar U, editor. J Qual Maint Eng. 2013 Aug 9;19(3):233–77.
- [5] Kelly A. Maintenance Strategy: Business-centred Maintenance. Butterworth-Heinemann; 2012. 276 p.
- [6] Wireman T. Benchmarking Best Practices in Maintenance Management. 2nd Revised edition edition. New York: Industrial Press Inc.,U.S.; 2010. 256 p.
- [7] Alex Hill & Terry Hill. Manufacturing Operations Strategy. 3rd ed. United Kingdom: Palgrave Macmillan; 2009.
- [8] Al- Turki U. A framework for strategic planning in maintenance. J Qual Maint Eng. 2011 May 31;17(2):150–62.
- [9] Velmurugan RS, Dhingra T. Maintenance strategy selection and its impact in maintenance function: A conceptual framework. Int J Oper Prod Manag. 2015 Dec 7;35(12):1622–61.
- [10] Pintelon L, Pinjala SK, Vereecke A. Evaluating the effectiveness of maintenance strategies. J Qual Maint Eng. 2006 Jan 1;12(1):7–20.
- [11] Waeyenbergh G, Pintelon L. A framework for maintenance concept development. Int J Prod Econ. 2002;77(3):299–313.
- [12] Tsang, A. Strategic dimensions of maintenance management. J Qual Maint Eng. 2002 Mar 1;8(1):7–39.
- [13] D.N.P. Murthy, A. Atrens, J.A. Eccleston. Strategic maintenance management. J Qual Maint Eng. 2002 Dec 1;8(4):287–305.
- [14] Singh PJ, Smith A, Sohal AS. Strategic supply chain management issues in the automotive industry: an Australian perspective. Int J Prod Res. 2005 Aug 15;43(16):3375–99.
- [15] Parida A, Kumar U. Maintenance performance measurement (MPM): issues and challenges. J Qual Maint Eng. 2006 Jul 1;12(3):239–51.
- [16] Pintelon L, Van Puyvelde F. Maintenance performance reporting systems: some experiences. J Qual Maint Eng. 1997 Mar 1;3(1):4–15.
- [17] Tsang, A. A strategic approach to managing maintenance performance. J Qual Maint Eng. 1998 Jun 1;4(2):87–94.
- [18] Berges L, Galar D, Stenström C. Qualitative and quantitative aspects of maintenance performance measurement: a data fusion approach. Int J Strateg Eng Asset Manag. 2013;1(3):238–252.
- [19] Handy C. Understanding Organisations 4th Fourth Edition. Fourth Edition edition. Penguin; 2005.
- [20] Schein EH, Schein P. Organizational Culture and Leadership. 5th Revised edition edition. Hoboken: John Wiley & Sons; 2017. 416 p.
- [21] Brown A. Organisational Culture. 2 edition. Harlow: Financial Times/ Prentice Hall; 1998. 336 p.
- [22] Rollinson DD. Organisational Behaviour and Analysis: An Integrated Approach. 4 edition. Harlow, England ; New York: Financial Times/ Prentice Hall; 2008. 784 p.
- [23] Smith ME. Changing an organisation's culture: correlates of success and failure. Leadersh Organ Dev J. 2003 Aug;24(5):249–61.

The Role of ISO 55000 Standard in Asset Integrity

I. Ihemegbulem¹, D. Baglee¹, A. Wheatley¹

¹ University of Sunderland, Sunderland, SR1 3SD, UK

ABSTRACT

Industries are on the edge of a new trend in asset efficiency improvement, with the current pressures that organizations experience in realizing optimal value from their assets. The challenge for organizations is the necessity of maintaining, and often increasing, operational effectiveness, revenue, and customer satisfaction, while at the same time reducing capital, operating, and support costs (Mitchell, 2002).

The International Standard for Organization (ISO 55000) was developed specifically for Asset Management (AM), which consists of a series of three components, ISO 55000, ISO 55001 and ISO 55002. ISO 55000 is not just about maintenance but about value creation. Assets allow organizations to achieve strategic objectives and meet stakeholder needs. Managing assets optimally has become crucial for organizations to stay competitive in today's global market. ISO 55000 requires changes in policy, processes and people thereby challenging the status quo and it leads to a whole new regime of asset integrity. Asset integrity requires organizations to comply with standards and various regulations. Adopting ISO 55000 will allow organization to align the way assets are managed and maintained, it improves return on investments by reducing costs, while supporting asset value without sacrificing organizational objectives.

Physical assets are often taken for granted and with increased reliability, the greater revenue generating potential they have. Managing assets today is more complex, yet in many organizations, the management of those assets is often unfocused and inappropriate. Organizations with a strong profit focus should look at ISO 55000 from the perspective of finding ways to leverage it into lower costs and increased outputs.

The aim of this paper is to classify the necessary elements in ISO 55000, which organizations need to adopt in order to develop a new approach to asset integrity, creating a clearer way to implement an asset management system within an organization.

Keywords: Maintenance, Asset Management (AM), and ISO 55000.

Corresponding author: Ibifuro Ihemegbulem (Ibifuro.ihemegbulem@research.sunderland.ac.uk)

1. INTRODUCTION

The global demand for asset performance is outpacing available capacity and infrastructure.

Organizations with aging systems and outdated style of business practices have limited solutions to address this problem. There are growing expectations by organizations for optimal performance from existing assets, because of the increased demand and standards by clients on what organizations should deliver on a daily basis.

The challenges of overcoming international competition and market globalization forces have placed enormous pressure on maintenance system in the quest to improve efficiency and reduce operational costs. These challenges have forced maintenance managers to adopt tools, methods, and concepts that could stimulate performance growth and minimize errors, and to utilize resources effectively toward making the organization a "world-class manufacturing" or a "highperformance manufacturing" plant. To achieve this, Assets require the support of people, services and resources of many types including operators, maintainers repair facilities, consumables, spare parts, documentation and training. Furthermore, to ensure that these

supports are provided and budgeted for, several types of analysis are need. With these industrial, technological, and economic pressures, there is the heightened need for collaboration between service providers and asset owners to efficiently manage the asset owner's assets more effectively, and to comply with asset management standards. It is therefore important to understand the underlying factors that are critical to the success of asset management service collaboration and partnerships.

2. ASSET AND ORGANISATIONAL SUSTENANCE

ISO 55000 define an asset as a 'thing, an item or an entity that has actual or potential value' to an organisation. Assets can be classes as production, operating equipment and structures. All assets are not equal in terms of value, some are more critical to achieving the objectives of the organization than others. While, certain assets present a higher degree of risk to business goals than others.

ISO 55000 questions whether all assets are managed, and views asset management as a coordinated set of activities to realize maximum value of all assets. Asset integrity can help to deliver real business value, organizations should support employees understanding of value and understanding of risk. The function of assets is to support the delivery of services to the organization. If an asset does not, contribute effectively to a service over time, it should not be held or used. Decisions about asset possession and retention should always be service driven.

Asset dependent organizations need to be continually educated and diligent about the importance of classifying assets in terms of the impact of asset failure on the organization. Asset criticality grading should be used to identify and classify failure consequences thereby allowing organizations to effectively allocate often-limited asset care resources.

Asset performance will progress if equipment is sustained in an appropriate condition. To attain this, essential maintenance must be carried out on a routine basis. Senior managers and all personnel must acknowledge maintenance as important to asset criticality; and therefore, managing asset effectively requires excellent asset knowledge.

2.1. Asset Knowledge

The operational knowledge of critical asset allows asset managers and management to make instant and accurate forecasts leading to thorough business decisions.

Management should be aware of the location of their assets, elements or parts that are essential to principal equipment. Management should keep track of changes, upgrades and regulatory compatibilities, it is advisable to list the date and type of the decisions.

2.2. Asset Criticality

Asset criticality defines the possible importance of asset failure and consequences to the overall business. This is achieved by assessing asset failure consequences against progressed criteria within business impact factors. The business impact factors of safety, quality, throughput and cost should be used for assessment.

ISO 55000 series requires that a strategic approach should be used to determine the criticality of assets and asset systems, and weight them appropriately when making decisions. It is important for organisation to determine assets with the most important function in the organisation. By checking the criticality of an asset, stakeholders can systematically arrive at an agreement about which assets are important to the business and why, thereby appropriately assigning resources for their care. Organization must have list of all assets to be manage and ensure Asset managers are managing portfolio of assets efficiently and effectively.

3. MANAGING ASSETS EFFECTIVELY

It is fundamental for organizations to efficiently manage their assets. This requires effective planning, acquisition, operation and disposal of the utmost assets to meet present and future service delivery demands. Responsible managers should incorporate asset management into their organizations. Organizations can achieve Asset integrity is by:

- Matching assets with service delivery needs.
- Managing assets within available resources and legal/technical requirements.
- Ensuring assets are fit for purpose and aligned with organisation policy objectives.
- Organisation adopting a lifecycle approach to planning asset investment and management decisions;
- Monitoring, evaluating and improving the performance of assets.
- Establishing responsibility and accountability for maintaining, and efficiently and effectively using the assets already in place;
- Ensuring informed decision making through adequate asset information, including asset condition, existing and alternative asset use and residual value.
- Identifying and managing the risks of asset ownership and operation.
- Setting up a continuous improvement and adaptive management approach to asset investment policies and practices.

4. THE IMPORTANCE OF ISO 55000 IN EFFECTIVE ASSET CONTINUANCE

The ISO 55000 series of standards stipulate the requirement for the implementation of physical asset management in asset intensive organization. It is designed to assure the organization and its stakeholders, regulators, insurance underwriters, and investors that the organization has a system in place to manage their assets in ways that deliver value aligned with the organization's objectives. In addition, ISO 55000 spans the entire life cycle of an asset: from design, to engineering, procurement, installation, start-up, operation, maintenance, restoration, decommissioning, and disposal; this process is popularly referred to as 'from cradle to grave'.

The ISO 55000 series of standards are divided into three parts;

Firstly, the **ISO 55000 Asset Management** which provides critical overview, concepts and terminology. Secondly, the **ISO 55001 Asset Management** which specifies the requirements for an effective Management System. It contains series of the "shall" clauses, stating what an organization needs to do to comply with the standard. ISO 55001 necessitate organization to set up a life cycle management plan.

ISO 55002 Asset Management; offers interpretation and guidance for such a system to be implemented. It contains clauses, which gives guidelines as to how the requirements of the corresponding clause of ISO 55001 should be implemented. The life cycle management plan should also include the risk associated with the specific asset and the consequences of this risk. The process of determining when asset will fail helps to determine the life cycle of the assets and how to manage the asset efficiently. To ensure reliability and productivity of assets, organizations should adopt the following structures of ISO 55000:

4.1. Organizational context

The external and internal drivers of the organization should be outlined, limitations and capacity to fulfill the objectives of company e.g. regulation, finance, organizational culture and environment and values of the organization. Stakeholder's impact to decision-making process should be considered. The scope of AM in the organization must be clear.

4.2. Leadership

Senior management should be included when setting up asset management leadership in an organization, a clear policy that supports the organizational strategic plan be established and reviewed. Committed leadership compels senior management to ensure continual improvement, the availability of resources, and Effective Asset Management Policy.

4.3. Planning

Organization must create a complete business plan before its operations and it must continue to update and revise it. The strictness of advance planning influences the difference between success and failure. Therefore, Organization should have a strategic way of dealing with risks, prospects and its AM objectives. In the quest of planning efficiently how organization will operate and maintain its assets, ISO 55000 outlines requirement that should be meet and consider. These integrated plans would address what will be done, when and who will do it, and how it will be undertaken and evaluated.

4.4. Support

Cooperation and collaboration with other departments will be required for effective management of assets. Information must be accessible, documented, controlled, communicated and auditable to back the process.

4.5. Operation

Personnel must be competent and trained to manage or operate an asset. Plans and implementation should be fed back into the asset management system, including any activities that are outsourced, and involves change management activities.

4.6. Performance evaluations

Organization should determine the techniques to monitor and measure its asset and how the data will be analyze, evaluated and validated. The report is used to assess the performance of an asset and risk management.

4.7. Improvements

Organization must check that nonconformities with respect to the assets are documented and evaluated. Corrective and preventive action must be in place for continual improvement of Assets and AM. Implementation of ISO 55000 provide many advantages some of which include the following:

- Diligence to standards and attracts investors/funding.
- Increase in asset financial returns
- Validates need for infrastructure investments
- Retains competent staff
- Reduces insurance rates.
- Manages asset and business risk
- Manages asset and business risk

5. CONCLUSION

Organizations need a robust risk management process to improve reliability, without it assets can struggle to perform and machinery may actually fail at the most inopportune time. It is important for organizations to know which assets are the most critical and where investment strategies should be placed more than the maintenance department. Organizations should stop owning collection of

assets that do not work well within their designated process or equipment that have hit the end of useful life and are due for replacement. The use of monitoring technology to monitor asset health is suggested (Lightfoot 2011). An advanced monitoring will result in higher asset availability.

Planning and scheduling activities for maintenance work should be conducted by classifications on a priority basis, the most critical should be chosen first for execution and the next is performed in turn until all back orders are completed.

ISO 55000 is not a maintenance strategy; companies must have a holistic maintenance strategy that will aid in improving the overall effectiveness, performance of their asset. Maintenance and reliability processes, programs, and best practices fit within ISO 55000, which puts strong emphasis on continual improvement and preventative action.

ISO 55000 provides opportunities to re-examine and improve asset ownership and service provider relationships, regulatory frameworks and stakeholder confidence. Therefore, ISO 55000 remains the best choice for organizations seeking competitive advantage and profitability through improved asset integrity.

REFERENCES

- [1] Cagliano, R. & Spina, G., 2000. Advanced Manufacturing Technologies and Strategically Flexible Production. *J. of Operations Management*. Volume 18, pp. 169-190.
- [2] ISO/CD55000, 2012. Asset management - Overview, principles and terminology. Lightfoot, H.W., Baines, T. and Smart, P. 2011. "Examining the information and communication technologies enabling servitized manufacture" *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 225 (10): 1964-1968.
- [3] Mitchell, J.S. 2002. *Physical Asset Management Handbook*. 3rd ed. Houston, TX: Clarion Technical.
- [4] Saranga, H. and J, Knezevic. 2000. "Reliability prediction for condition based maintained systems" *Reliability Engineering and System Safety*, 71 (2): 219-224.
- [5] Woodhouse, J., 2012. *ISO 55000 Standards for asset management*.

Session 6

Monitoring, Diagnosis, Prognosis and Health Management 2

Session Chair Asst. Prof. Hugo Andre

Incipient Fault Diagnosis and Monitoring System for Planetary Gear Train

(L. Gao, L. Wang, Y. Shao, L. Yin, M. Du, Y. Yang, C. Zheng)

Spline Wavelet Based Filtering for Denoising Vibration Signals Generated by Rolling Element Bearings *(J. Kansanaho, K Saarinen, T. Kärkkäinen)*

Fault Diagnosis of Early-Stage Spalling Fault Based on a Hybrid Approach

(L. Wang, Y. Shao, M. Du, Y. Yang, F. Guo, Z. Cao)

Bogie Speed Estimation and Signal Source Separation via Rail Vibration Analysis

(F. Gu, Y. Xu, F. Liu, A. D. Ball)

Incipient Fault Diagnosis and Monitoring System for Planetary Gear Train

Linghan Gao¹, Liming Wang¹, Yimin Shao¹, Lei Yin¹, Minggang Du², Yang Yang², Cao Zheng¹

¹ State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing, CHINA

² China North Vehicle Research Institute, Beijing, CHINA

ABSTRACT

Planetary gear train are widely used in wind turbines on account of their properties of lower weight, higher transmission ratio and efficiency, bigger bearing capacity in comparison with fixed-shaft gear train. In wind turbine, the failure rate of planetary gearboxes is extremely high due to the harsh operation conditions. These gearboxes are often installed far off the ground which makes the maintenance cost very high. Due to these facts, developing the techniques for gearbox condition monitoring and early fault diagnosis really makes a lot of sense. Aimed at fault features of planetary gear train, the On-line monitoring system based on LABVIEW can achieve vibration monitoring, rotation monitoring, temperature monitoring and historical data query under a variety of operating conditions. To capture the operational status and to evaluate the health status of the wind turbine in real time, and to transmit it synchronously to the network, so as to provide theoretical guidance and basis for wind turbine maintenance.

Keywords: monitoring system, wind turbine, gearbox.

Corresponding author: Prof. Yimin Shao (ymshao@cqu.edu.cn)

1. INTRODUCTION

Wind energy has attracted more and more attentions in the past decades, because of the global climate deterioration and traditional fossil energy reduction. According to the Global Wind Energy Council (GWEC), the global totally installed wind power had reached more than 487GW, nearly half of which is added in the past five years [1].

Wind turbines are expensive and the scheduled maintenance fee is very high, so that condition-based maintenance (CBM) is needed, which can help reduce the maintenance fee and improve the reliability of the wind turbine. Wind turbines fault statistics of Sweden, Finland and Germany [2] are listed in table 1.

From the statistics, we can know that the number of electrical system of wind turbines is the largest. However, gearbox fault cause the maximum downtime of wind turbines. Among the common gear faults, tooth crack is usually observed at the tooth root because of excessive loads transmitted and the material fatigue especially [3, 4]. Once the gearbox has failed, it is difficult to repair on the tower due to the cabin space limited. And the whole maintenance cycle is very long, which will seriously affect the economic benefits. It is meaningful to install the on-line monitoring system on the wind turbine gearbox, which can monitor the operation status in real time and timely discover the hidden fault.

Gearbox with one-stage planetary and two parallel shaft is widely used in wind turbines. In this study, a diagnosis and monitoring system is designed to collect vibration signal, temperature signal, shaft displacement and rotating speed of gearbox. And then analyze data, evaluate health status and transmit it synchronously to the network.

Table 1: Wind turbines fault statistics of Sweden, Finland and Germany

| | Sweden | Finland | Germany |
|-------------------------------------|----------------------------------------------------------|--------------------------------------------|-------------------------------------------------------------------------------------|
| Annual average number of faults | 0.402 | 1.380 | 2.380 |
| Annual average downtime (H) | 52 | 237 | 149 |
| Average downtime of single fault(H) | 170 | 172 | 62.6 |
| The most faulty system | 1. electrical system 2. sensor of control 3. blade | 1. hydraulic system 2. blade 3. gear | 1. Electrical system 2. Control system 3. Hydraulic system, sensor of control |
| Maximum downtime failure parts | 1. Gear 2. Control system 3. Electrical system | 1. Gear 2. Blade 3. Hydraulic system | 1. Generator 2. Gear 3. Transmission chain |
| Single failure downtime ranking | 1. Transmission chain 2. Yaw system 3. Gear | 1. Gear 2. Blade 3. Structure | 1. Generator 2. Gear 3. transmission |

2. SELECTION OF EVALUATION INDEX OF OPERATING STATE OF A WIND TURBINE GEARBOX

2.1. Vibration

Using Fast Fourier transform to vibration signal, we can get the frequency spectrum. If the gear fails, its vibration signal is often expressed as the rotation frequency, tooth meshing frequency and its frequency modulation. On the spectrum, there are two equally spaced sidebands centred on the meshing frequency. The sidebands reflect the source of the fault. The interval of the sideband reflects the frequency of the fault. The change of the amplitude reflects the severity of the failure. So the vibration is the primary parameters of evaluation index of operating state of wind turbine gearbox.

Accelerometer sensors are arranged in the main bearing housing, gearbox input and output. Monitoring system can display and analyse vibration value. If the value is higher than threshold, the system will perform a dual alarm for the indicator and buzzer.

2.2. Displacement

The shaft is a part that transmits movement and power. Main shaft of wind turbine will bear bending moment and torque moment. When the wind speed is too large, it may cause the bending, radial or axial displacement of main shaft. Displacement sensor is arranged in main bearing housing.

2.3. Rotating speed

The rotating speed of shaft of gearbox will fluctuate in a large range due to the size of the wind. That monitoring the rotating speed of running gearbox is conducive to help analyse the vibration value, displacement value and temperature.

2.4. Bearing oil temperature

The bearing is a part that supports and guides the movement of the shaft, aimed at reducing the friction coefficient of the shaft during transmission. When the wind turbine gearbox is running, friction will produce a certain amount of heat. When the heat balance, the temperature will be fixed in a normal value. If there are defect, heat balance would be broken. The temperature will rise sharply. If the value of temperature is higher than threshold, the system will perform a dual alarm for the indicator and buzzer.

3. MONITORING SYSTEM DESIGN

3.1. Measuring point arrangement

Design for one main bearing, gearbox with one-stage planetary and two parallel shaft. Measuring point arrangement is shown in figure 1. Sensor parameters is listed in Table 2

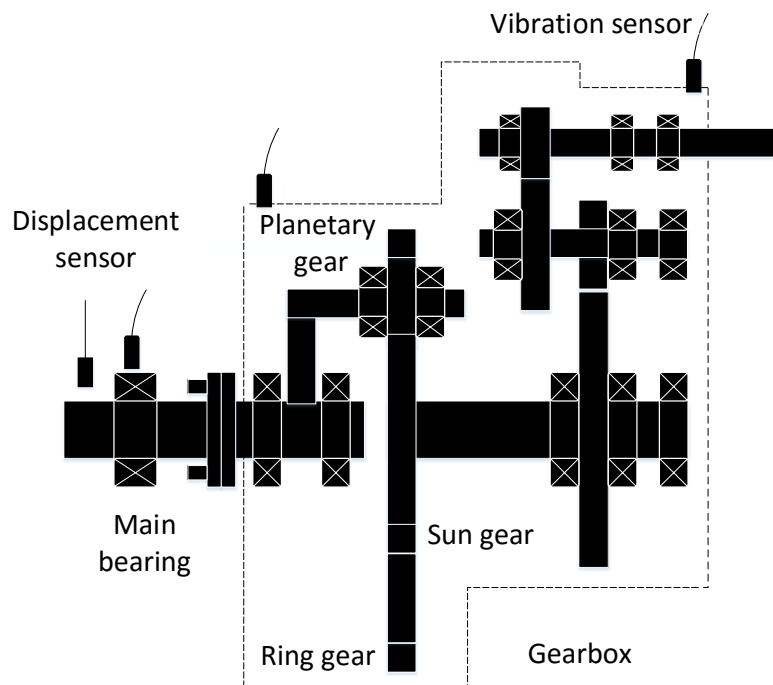


Figure 1: Measuring point arrangement

Table 2: Sensor parameters

| Sensor | Range | Reference sensitivity | Frequency range |
|--------------|-----------|-----------------------|-----------------|
| vibration | $\pm 50g$ | 100 mv/g | 0.5~10k Hz |
| displacement | 4mm | 4 v/mm | |

3.2. Monitoring system hardware design

System hardware design is shown in figure 2. Including: sensors, data acquisition card, fiber optic transceiver, fiber optic cable, router, industrial PC.

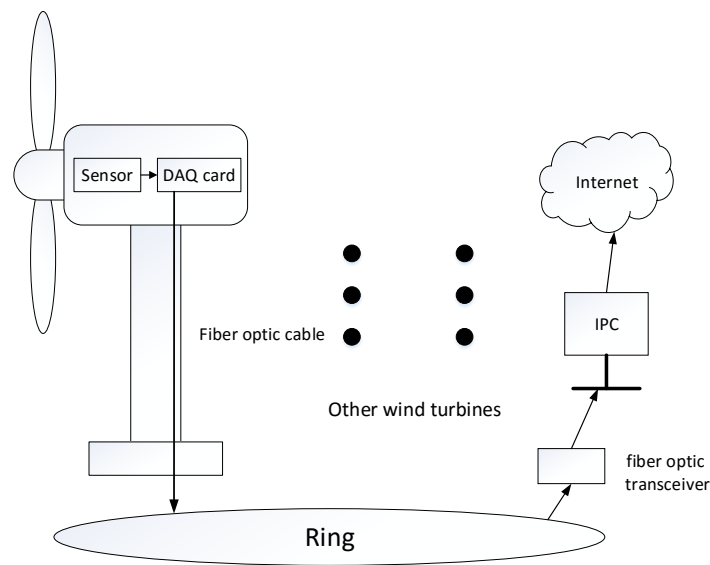


Figure 2: Monitoring system hardware design

3.3. Monitoring system software design

In this study, the monitoring system is based on LABVIEW, which is a professional monitoring and control language development platform. The main functions of the monitoring system software include multichannel continuous data acquisition, parameter settings, waveform display, spectrum analysis, early warning function and data storage. Multichannel settings is shown in figure 3. Waveform display is shown in figure 4. Spectrum analysis is shown in figure 5. The RMS value of the signal reflects the energy of the signal. It's an indicator of the strength of the dynamic signal. If gearbox have defect, the value of RMS will increase. The system issues an alert.

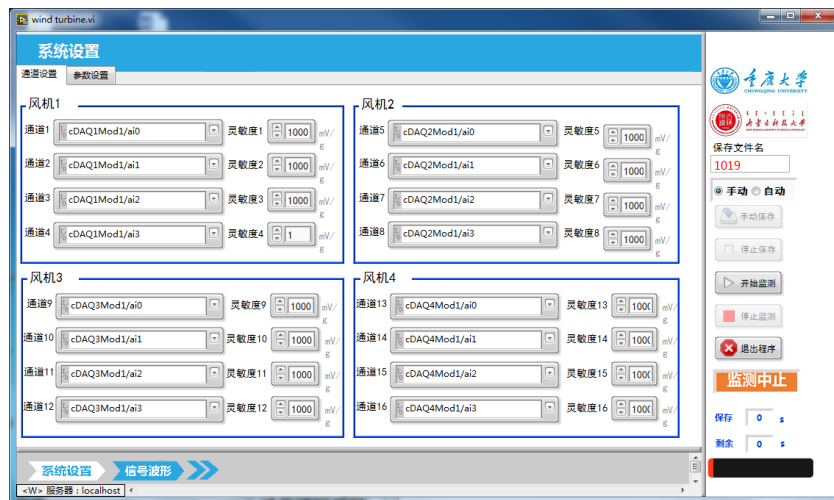


Figure 3: Multichannel settings

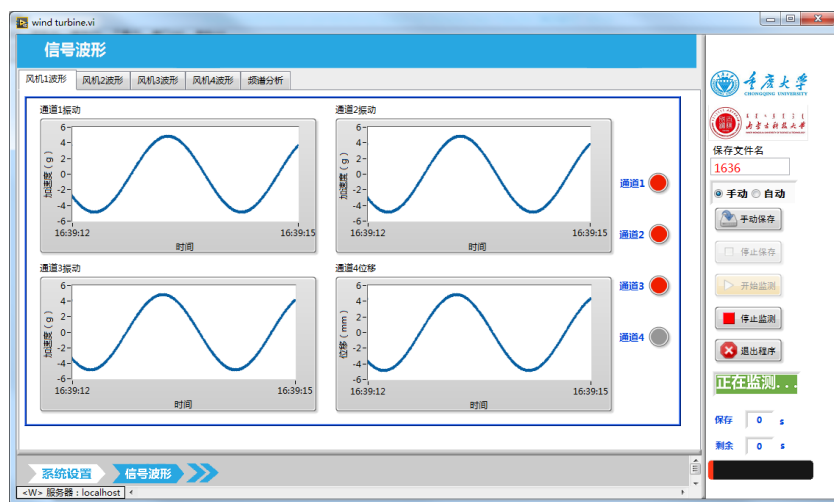


Figure 4: Waveform display

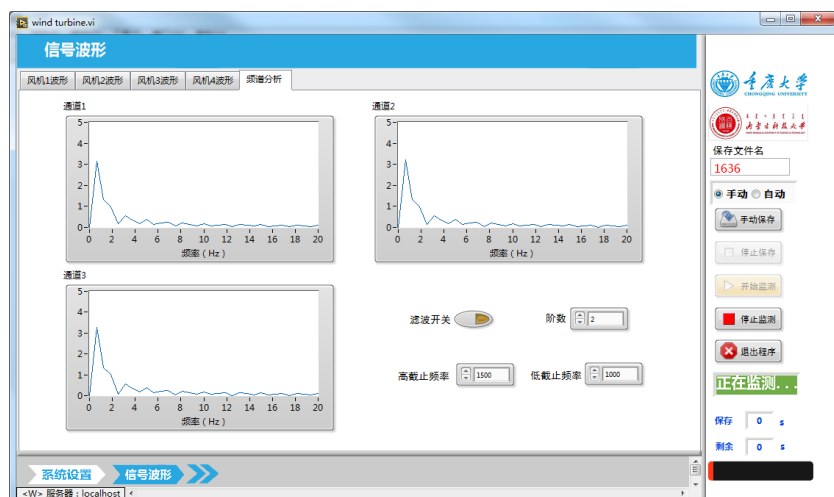


Figure 5: Spectrum analysis

4. CONCLUSIONS

This study presented a monitoring system of wind turbine gearbox. It can achieve vibration monitoring, temperature monitoring, rotating speed monitoring and displacement of shaft monitoring. On-line monitoring system can timely identify the operating state of the gearbox. It can discover the feature of the incipient fault, to provide theoretical guidance and basis for wind turbine maintenance.

ACKNOWLEDGEMENT

The authors are grateful for the financial support provided by the China North Vehicle Research Institute and the National Natural Science Foundation of China under Contract No.51475053.

REFERENCES

- [1] Sheng S, Link H, et al. Wind turbine drivetrain condition monitoring during GRC phase 1 and phase 2 testing. *Contract*, 2011, 303: 275-3000.
- [2] Ribrant J. Reliability Performance and Maintenance-A Survey of failures in Wind PowerSystems [D]. Sweden: KTH School of Electrical Engineering, 2006.
- [3] Chaari F, Fakhfakh T, Haddar M. Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness. *European Journal of Mechanics-A/Solids*, 2009, 28(3): 461-468.
- [4] Chaari F, Baccar W, Abbes M S, et al. Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. *European Journal of Mechanics-A/Solids*, 2008, 27(4): 691-705.

Spline Wavelet Based Filtering for Denoising Vibration Signals Generated by Rolling Element Bearings

J. Kansanaho¹, K. Saarinen² and T. Kärkkäinen³

¹ University of Jyväskylä, Department of Mathematical Information Technology, PO Box 35, FI-40014 University of Jyväskylä, Finland; jarno.m.kansanaho@jyu.fi; phone +358400248111

² ABB AB, Corporate Research, Forskargränd 8, SE-721 78, Västerås, Sweden; kari.saarinen@se.abb.com

³ University of Jyväskylä, Department of Mathematical Information Technology, PO Box 35, FI-40014 University of Jyväskylä, Finland; tommy.karkkainen@jyu.fi

ABSTRACT

In this paper, we study spline wavelet filtering methods to denoise vibration signals generated by rolling element bearings. Common challenges in the analysis of vibration measurements taken from a real industrial environment is that non-stationary components, generated by other machine components, disturb the analysis. Vibration signals generated by non-faulty and faulty rolling element bearings are studied. As known, the Fourier transformation does not work very well on non-stationary signals because their spectral content changes over time. In the time-frequency domain methods signal decomposition is performed to split spectrum into sequential sub-spectral components that are processed individually. The weakness of the short-time Fourier transform is that the constant window size does not provide sufficient frequency and time resolution at the same time. Lately, the wavelet transform has been applied on signal demodulation and optimal band-pass filter design. More flexible than basic wavelet basis are spline wavelets that are constructed with a spline function. Spline wavelets are linear combination of B-splines and they can be defined explicitly. In our study spline wavelet based denoising is applied on computer simulated vibration signals and vibration signals acquired from a real-world application.

Keywords: Rolling element bearings, Vibration analysis, Advanced signal processing, Spline wavelets

Corresponding author: Jarno Kansanaho (jarno.m.kansanaho@jyu.fi)

1. INTRODUCTION

Rolling element bearing (REB) is one of the most common component in various machines, motors, wheels, cooling fans etc., to support rotating shafts. Failures of bearings are the most common reason for machine breakdowns. Sudden breakdowns can be very expensive. Consequently, detection of the bearing fault as early as possible is very valuable. An important part of the machine condition monitoring is to detect weak impacts generated by faulty REBs. Certain sensors are used to detect the vibrations of these impacts. However, in the noisy industrial environment other machine components generate vibrations that disturb the analysis.

The main motivation for this work is that the weak bearing fault signatures experience interference from noise from different external sources and internal mechanical components, which makes bearing fault detection from real measurements a very challenging task. De-noising (noise reduction) is the key technique to reduce the effect of the disturbing components in the analysis.

As known, the Fourier transformation does not work very well on vibration signals that contain short and high frequency pulses, which very important in nonstationary signals [1]. In the time-frequency domain methods, the signal decomposition is performed to split the spectrum into sequential sub-spectral

components that are processed individually. The sinusoidal and phase information of local sections of the signal are determined by Short-Time Fourier transform (STFT). Weakness of the STFT is that the constant window size does not provide sufficient frequency and time resolution at the same time [2].

Lately, the wavelet transform has been applied for signal demodulation and optimal band-pass filter design [3]. The wavelet transform has been successfully applied to bearing fault detection [4, 5]. However, there is no standard method to select the wavelet function for different purposes, so that the wavelets do not have a standard status in fault diagnostics [6]. Breakdown points, trends and discontinuities in higher derivatives are detectable by using wavelet analysis [7]. Verma and Sreejith introduced a Morlet wavelet based filtering method that efficiently detects weak bearing fault impulses even the signal to noise ratio is very low [8]. Yumona et. al. used a Morlet wavelet bank for denoising and obtained the resonance band of interest from the wavelet kurtogram [9]. The wavelet packet transform has been utilized often in the condition monitoring of the rolling element bearings [10, 11, 12]. It has been noticed that the Wavelet packet transform (WPTT) has better de-noising ability on non-stationary signals because the frequency resolution of the discrete wavelet transform (DWT) may not be enough to extract important features from the decomposed part of the signal [10].

2. THEORETICAL BACKGROUND

The wavelet transform enables multiresolution analysis with dilated windows. In other words, the wavelet transform is a constant relative bandwidth analysis [13]. By using dilation and translation, the wavelet transform can extract time-frequency features of a signal efficiently [9]. It is an excellent tool to analyse nonstationary signals. The frequency resolution increases and the time resolution decreases when the wavelet transform is done at sequentially wide scales. Like Fourier transform the wavelet-transform concludes the basis-function called wavelet.

The continuous wavelet transform (CWT) transforms signal to a two-dimensional time-scale joint representation. The idea of the CWT is to calculate continuously scalable function by moving this function continuously over a signal. As result, the wavelet coefficients are acquired. However, the bases of the scalable functions become non-orthogonal that makes wavelet coefficients redundant [14]. The definition of the CWT as the function of time is:

$$T(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{(a,b)}^*(t) dt, \quad (1)$$

where ψ is the mother wavelet and $*$ refers to complex conjugation. The definition of a mother wavelet is:

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (a, b \in R, a \neq 0), \quad (2)$$

where a is the scaling (dilation) parameter, and b if the translation parameter. The parameter a controls the window length and effects the frequency resolution; large a for better frequency resolution.

The time-scale joint representation of a discrete wavelet transform (DWT) is a grid along the scale and time axes. The discrete wavelet is a piecewise continuous function. The discretization of the wavelet is done by sampling the time-scale axis at discrete intervals. Usually dyadic sampling is used; a geometric sequence of ratio 2. The DWT as a function of time:

$$\psi_{(i,j)}(t) = \frac{1}{\sqrt{2^i}} \left(\frac{t-2^i j}{2^i} \right), \quad (3)$$

where the dilation term is 2^i and translation term is $2^i j$.

As one weakness, the wavelet transform does not maintain the absolute phase of the signal components. A wavelet is orthogonal when the corresponding wavelet transform is orthogonal. When the wavelet transform is invertible the corresponding wavelet is biorthogonal (not necessarily orthogonal). The number of degrees

of freedom increases when biorthogonal wavelets are used. For biorthogonal wavelets there are two scaling functions and the scaling sequences may differ. Designing biorthogonal wavelets allows more degrees of freedom than orthogonal wavelets. The semi-orthogonal wavelets were introduced when relaxation of the intra-scale orthogonality constraint was founded [15]. Wavelets are categorized into families based on their properties. Some of the important properties of the wavelets are regularity, symmetry or anti-symmetry, number of vanishing moments and existence of a scaling function. In practice, for example a greater number of vanishing moments provide sharper frequency resolution.

More flexible wavelets than basic wavelet basis are spline wavelets that are constructed with a spline function. Spline wavelets are categorized to interpolatory spline wavelets, B-spline wavelets, cardinal B-spline wavelets, and Battle-Lemarie wavelets. The polynomial and discrete splines are a source for a family of filters, which generate biorthogonal wavelets [16]. The greatest benefits of using spline wavelets are [17]:

- Polynomial spline bases have a simple and explicit analytic form that is easy to manipulate; differentiation and integration.
- The B-splines have compact support.
- Any degree of regularity is achieved by increasing the order of the polynomial splines.
- Polynomial splines are piecewise constant functions in the simplest case.

A polynomial spline function of degree n is defined by a linear combination of shifted B-splines [17]:

$$g^n(x) = \sum_{k=-\infty}^{+\infty} c(k)\beta^n(x-k), \quad (4)$$

where $c(k)$ is the sequence of B-spline coefficients and $\beta^n(x)$ is the central B-spline of order n which definition is:

$$\beta^n(x) = \beta^n * \beta^{n-1}(x), \quad (5)$$

where $\beta^0(x)$ is the characteristics function and $*$ is the convolution operator. B-spline is visualized in the Figure 1.

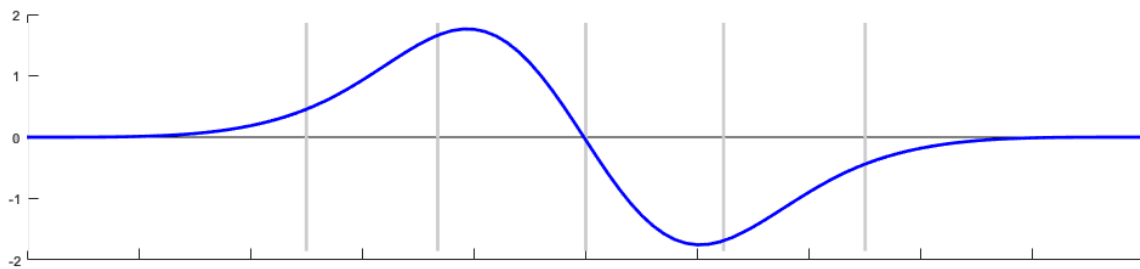


Figure 1 B-spline of order 6 contains six polynomial pieces of order 5

The compactly supported B-spline wavelet of order m is defined by the equation [18]:

$$\psi_m(x) = \frac{1}{2^{m-1}} \sum_{j=0}^{2m-2} (-1)^j N_{2m}(j+1) N_{2m}^{(m)}(2x-j), \quad (6)$$

when m is set to 1, the equation 5 defines the Haar wavelet [18]:

$$\psi_1(x) = \begin{cases} 1, & \text{when } 0 < x < 1/2 \\ -1, & \text{when } 1/2 < x < 1 \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Biorthogonal spline wavelets are regular, compactly supported and have FIR (finite impulse response) implementation. Biorthogonal spline wavelet filtering is applied to de-noising of the vibration signals in the

experimental part of this work. The construction of wavelet filter banks concludes the analysis and synthesis functions that perform composition of the original spectrum using sub-spectral components. Figure 2 incorporates decomposing and reconstructing scaling and wavelet functions of the biorthogonal spline wavelet (bior6.8).

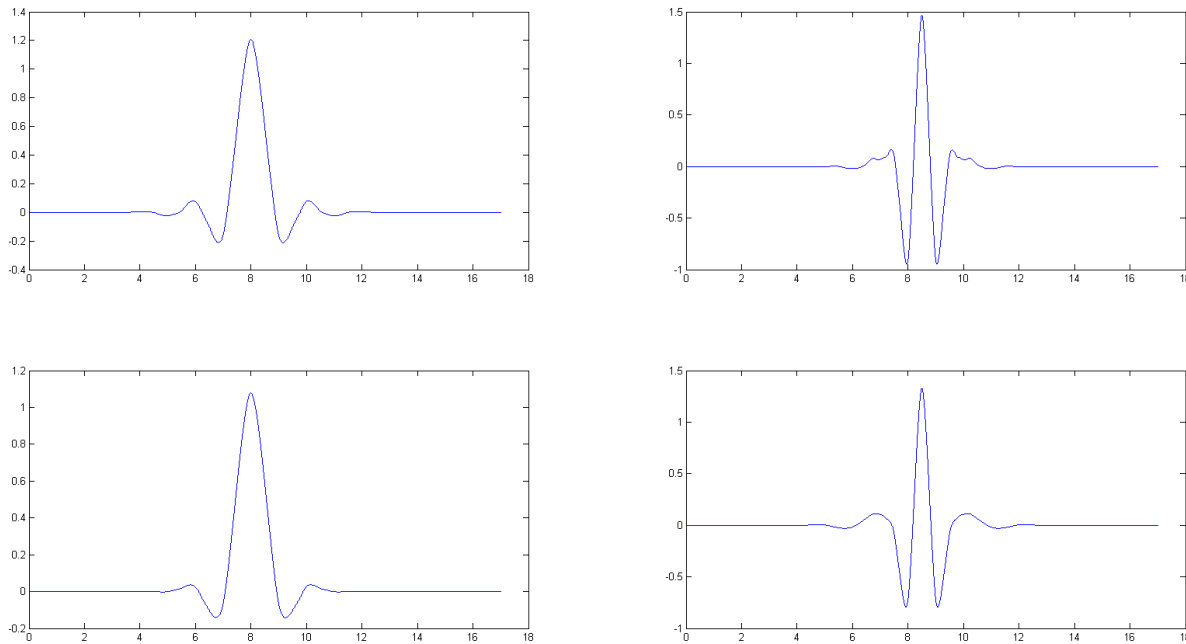


Figure 2 Decomposition and reconstructing wavelets and scaling functions of the bior6.8

3. EXPERIMENTAL PART AND RESULTS

Spline wavelet based filtering for denoising of vibration signals is presented in this case study. Simulated and real-world measured vibration signals are analyzed. The process of the experimental work is the following:

- Vibration signal preparation
 - Simulation of background noise, disturbing pulses and bearing fault pulses
 - Import measured vibration signal data
- The CWT study of the vibration signal using spline wavelets
- Vibration signal denoising using spline wavelet filtering with soft thresholding
- Study of the original (FIR filtered) and the denoised signal by using envelope analysis

The first case incorporates the analysis of the simulated vibration signal. The vibration signal model presented by Kansanaho et.al. is applied in the simulation [19]. The resonant frequency of the bearing fault pulse is 500 Hz and the bearing fault frequency is 11.0 Hz. The disturbance pulse is modeled using wavelet function and its repetition frequency is 5.0 Hz. Addition to, white noise is added to signal by signal to noise ratio of 5.0. The top most image in Figure 3 incorporate the vibration signal of the first case. The CWT is performed on the vibration signal by using biorthogonal spline wavelet (bior1.3). The bottom image in Figure 3 show the scaleogram of the signal. The bearing fault pulses are clearly extracted as seen in the figure 3. The vibration signal is high-pass filtered with FIR-filter (1500 Hz cut-off-frequency) before the envelope analysis. The Figure 4 show the envelope spectra of the FIR-filtered signal and the wavelet filtered signal. The frequency axis of the envelope spectrum is normalized according to the rotation speed. The envelope spectrum of the FIR-filtered vibration signal includes the both disturbance and bearing fault frequencies and their harmonics (left in the Figure 4). The wavelet-filtered signal does not contain the low

frequency disturbance and the bearing fault frequency and its harmonics are clearly seen (right in the Figure 4).

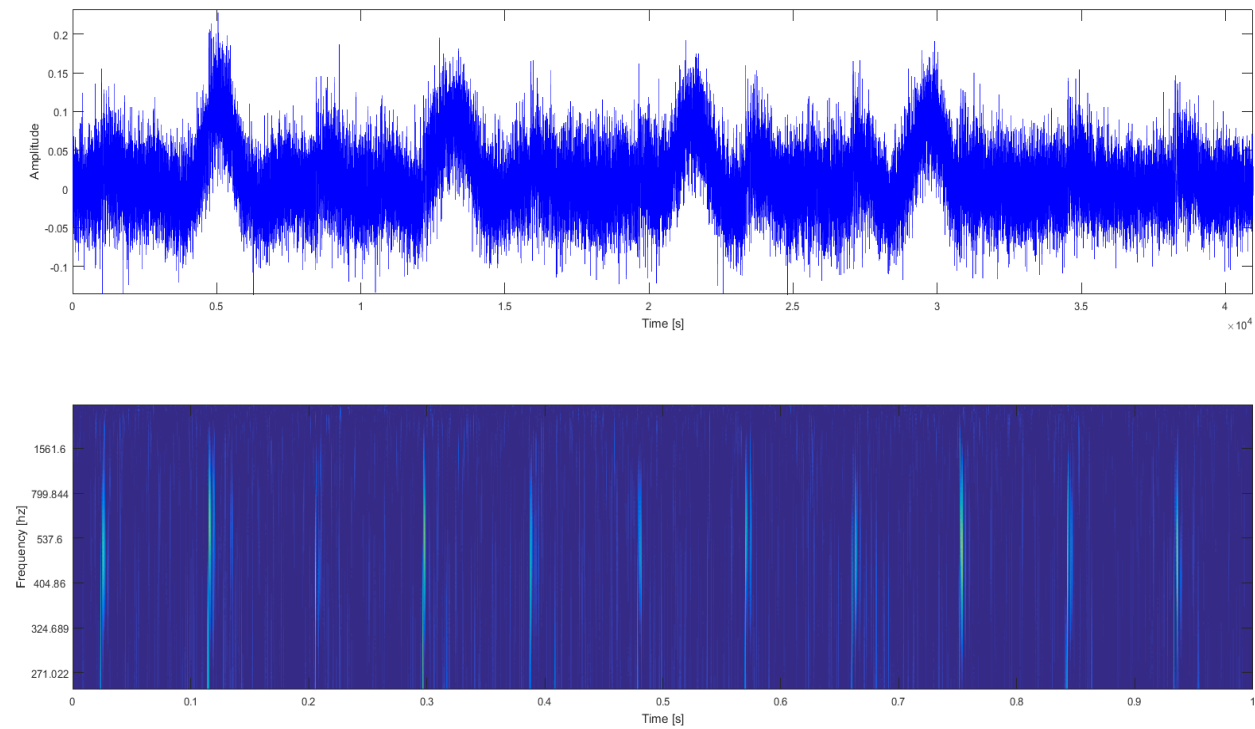


Figure 3 Case 1: The simulated vibration signal (top) and the CWT analysis of the signal (bottom).

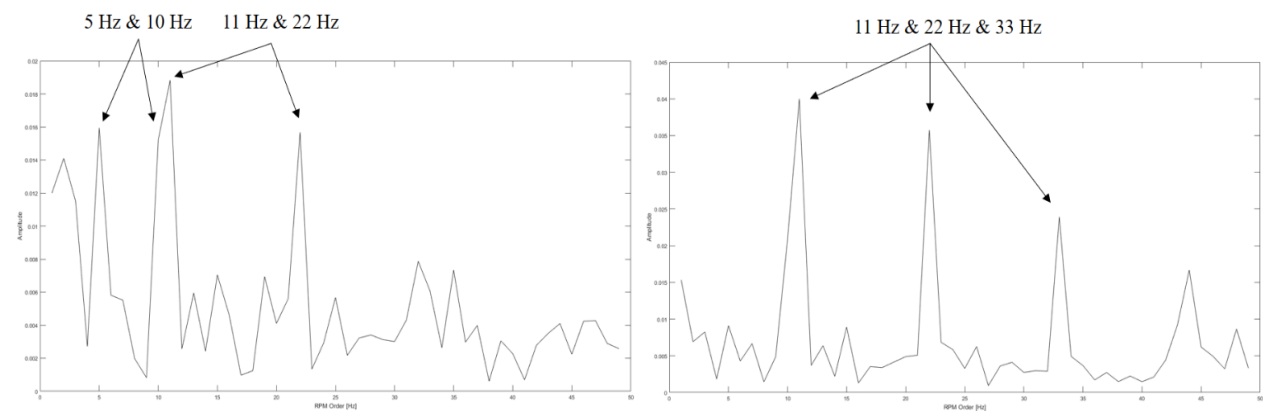


Figure 4 Case 1: The envelope spectrum of the FIR-filtered (high-pass) signal (right) and the wavelet-filtered vibration signal (left).

A noise-rich measured vibration signal is analyzed in the second case. The figure 5 incorporate a certain part of the signal (top) and the scaleogram of that part of the signal (bottom). Biorthogonal spline wavelet (bior6.8) is applied in the analysis. The resonant frequency (200 Hz) of the pulses is related to the bearing fault pulses 105 Hz (BPFI) as shown in the Figure 5 (bottom). The inner race bearing fault frequency and its second harmonic are marked with dotted red lines into the Figure 6, which incorporate the envelope spectra of the signals. It is hard to make conclusion that there is an inner race fault in the bearing based on the envelope spectrum of the FIR-filtered signal. The envelope spectrum of the wavelet-filtered signal reveals the peaks more clearly. However, still making a conclusion of the bearing fault is a little uncertain due the noise level in the envelope spectrum.

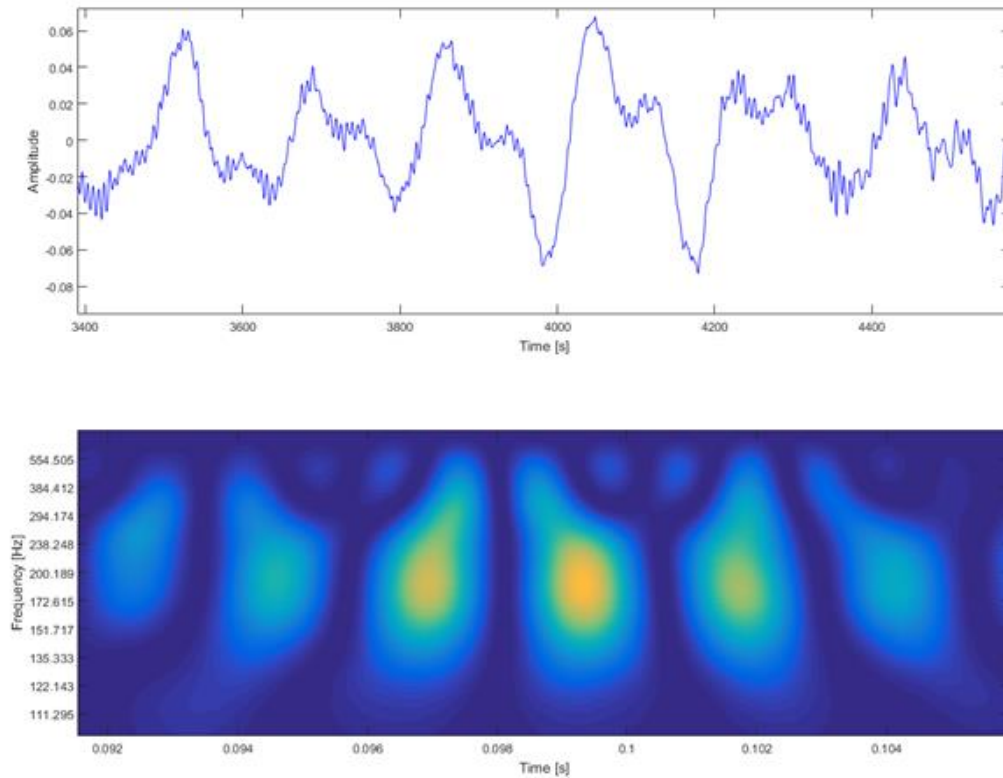


Figure 5 Case 2: A part of a measured noisy vibration signal (top) and the CWT analysis (scaleogram)

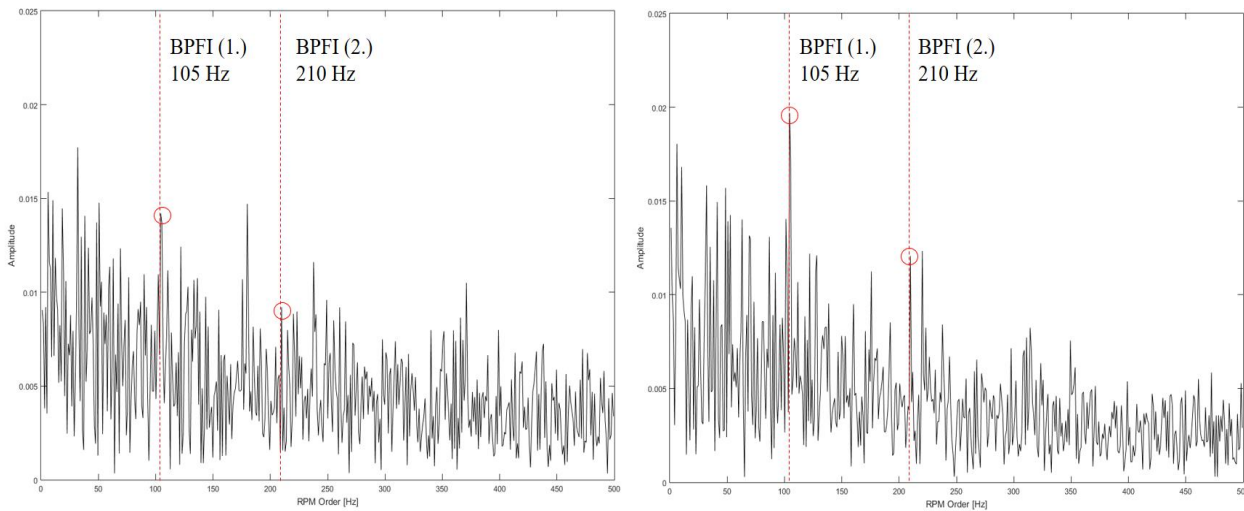


Figure 6 Case 2: The envelope spectrum of the FIR-filtered signal (right) and the wavelet-filtered signal (left).

4. CONCLUSION AND DISCUSSION

Spline wavelet denoising on presented vibration signals shows promising results. Non-stationary components of the vibration signal can be extracted and if necessary, they can be filtered. The filtering process itself was not automatic; the selection of the filtering coefficients done was manually by interpreting the wavelet decomposition of the vibration signals. In the future, the process could be automated. The flexible simulation of the vibration signals is a great tool when developing and testing denoising filters. There is a lot of research and practical applications done with the wavelet denoising. Such

as the beginning was said, that there is no standard method to select the wavelet function for different purposes. The spline wavelets are flexible and the implementation to fast algorithm is possible. Further studies include exploitation of the wavelet packets and frames.

REFERENCES

- [1] Sejdic, E., Djurovic, I., and Jiang, J. (2009). Time–frequency feature representation using energy concentration: An over-view of recent advances Digital Signal Processing, 19, 153–183
- [2] James Li, C. and Ma, J. (1997). Wavelet decomposition of vibrations for detection of bearing-localized defects NDT&E International, Vol. 30, No. 3, 143-149
- [3] H. Qiu, J. Lee, J. Lin and, G. Yuc (2006), Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, Journal of Sound and Vibration 289, 1066–1090
- [4] Rubini, R. (2001). Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings, Mechanical Systems and Signal Processing, 15(2), 287-302
- [5] Luo, G. Y., Osypiw, D., and Irle, M. (2003). On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing, Journal of Vibration and Control, 9, 931-947
- [6] Peng, Z. K. and Chu, F. L. (2004). Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing, 18,199–221
- [7] Kumar, H.S., Srinivasa Pai, P., Vijay, G.S. & Raj B.K.N. Rao (2014). Wavelet Transform for Bearing Condition Monitoring and Fault Diagnosis: A Review, 17(1), 9-23
- [8] Verma, A. K. and Sreejith, B. (2009), Rolling element bearing fault diagnosis using adaptive Morlet wavelet filter, International Journal of COMADEM, 12(4), 25-32
- [9] M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B. G. Celler, S. W. Su (2016), Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model, Engineering Applications of Artificial Intelligence 47, 88-100
- [10] J. L. F. Chacon, V. Kappatos, W. Balachandran, and T. H. Gan (2015), A novel approach for incipient defect detection in rolling bearings using acoustic emission technique. Applied Acoustics 89, pp. 88–100
- [11] F. Slouki, M. El Aroussi, H. Medromi, M. Wahbi (2013), Bearings prognostic using mixture of Gaussians hidden Markov model and support vector machine, Proceedings of IEEE ACS international conference on computer systems and applications (AICCSA), p. 1–4.
- [12] N. G. Nikolaou and I. A. Antoniadis (2002), Rolling element bearing fault diagnosis using wavelet packets, NDT&E International 35, pp. 197-205
- [13] J. Kovacevic and M. Vetterli (1995), Wavelets and Subband Coding, Prentice Hall PTR
- [14] Y. Sheng (2000), “Wavelet Transform.”, The Transforms and Applications Handbook: Second Edition, Ed. Alexander D. Poularikas, Boca Raton: CRC Press LLC
- [15] M. Unser, P. Thenenaz, and Akram Aldroubi (1996), Shift-Orthogonal Wavelet Bases Using Splines, IEEE Signal Processing Letters, Vol. 3, No. 3
- [16] A. Z. Averbuch, P. Neittaanmäki, and V. A. Zheludev (2014), Spline and Spline Wavelet Methods with Applications to Signal and Image Processing, Volume I: Periodic Splines
- [17] M. Unser, P. Thenenaz, and Akram Aldroubi (1992), “Polynomial Splines and Wavelets – A Signal Processing Perspective”, Wavelets – A tutorial in Theory and Applications, Ed. C. K. Chui, pp. 91-122
- [18] C. K. Chui and J.-Z. Wang (1992), On Compactly Supported Spline Wavelets and a Duality Principle, Transactions of the American Mathematical Society, Volume 330, Issue 2, pp. 903-915
- [19] J. Kansanaho, K. Saarinen, and T. Karkkainen (2016), Flexible Simulator for the Vibration Analysis of Rolling Element Bearings, COMADEM 2016 - the 29th International Congress on Condition Monitoring and Diagnostic Engineering Management

Fault Diagnosis of Early-Stage Spalling Fault Based on a Hybrid Approach

Liming Wang¹, Yimin Shao¹, Minggang Du², Yang Yang², Fang Guo¹, Zheng Cao¹

¹ State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing, CHINA

² China North Vehicle Research Institute, Beijing, CHINA

ABSTRACT

Due to heavy load and tough working environment, spalling faults are usually observed on gear tooth, which have a bad influence on vibration characteristics of the gear system. However, it is still an unsolved puzzle to detect the early stage gear fault. In this paper, a hybrid approach for fault diagnosis of spalling faults is developed by employing the technique of time synchronous averaging (TSA), spectral entropy (SE) and envelope analysis (EA). The interference of environmental noise and other components is suppressed by TSA firstly, then the demodulation band for EA is selected based on an optimized SE technique. The effectiveness of the proposed approach is validated by analyzing experimental signals from a two-stage gearbox with a seeded spalling fault.

Keywords: Fault diagnosis; spalling; TSA; spectral entropy; envelope analysis.

Corresponding author: Prof. Yimin Shao (ymshao@cqu.edu.cn)

1. INTRODUCTION

Gearboxes are widely used in wind turbines, helicopters, automobiles and industrial equipment to transmit power and produce high rotational speed changes and change the direction of motion. However, due to tough working environment or improper gear design, unexpected gear failures may occur, which reduce the reliability of the whole equipment and may cause large economic losses, even catastrophes [1, 2]. Thus, research on effective diagnosis of early-stage gear faults makes a great sense.

However, it is usually difficult to detect the early-stage gear faults, because the early-stage gear faults yield weak fault features which are easily immersed in background noise and normal gear meshing signals [3]. Therefore, research on signal de-noising has attracted a lot of attention from scholars around the world. For example, Tabrizi proposed a combined method to detect the early bearing faults, wavelet packet decomposition (WPD) is utilized for de-noising of the acquired vibration signal [4]. Zhao presented a reweighted singular value decomposition (RSVD) for signal de-noising and weak feature enhancement, and it is successfully used in the fault diagnosis for a two-stage gearbox as well as train bearings [5]. Braun revisited the signal processing tool of time synchronous averaging technique, the capability of TSA on periodic components extraction and noise suppression are discussed in detail [6]. Among the above methods, there are some parameters or threshold to be determined before the use of WPD or RSVD which may increase the difficulty for engineers without too much fundamental knowledge of signal processing. Therefore, TSA is adopted in this study for noise reduction.

As a commonly used method, envelope analysis method is commonly used in fault characteristic frequency detection for rotating machinery, i.e. bearings and gearboxes [7]. However, most times it is difficult to select the band-pass filter parameters, therefore, spectral entropy (SE) [8] is adopted in this study as a criterion to help select the optimal sub-band for envelope analysis. Hence, a hybrid approach based on TSA, SE and EA is established in this paper for early-stage gear fault diagnosis, in which interference of environmental noise is suppressed by TSA firstly, then the demodulation band for EA is selected based on the optimization of SE.

2. METHODOLOGY FOR EARLY-STAGE SPALLING FAULT DIAGNOSIS

2.1. Time synchronous averaging (TSA)

TSA refers to a method of averaging a large number of periodic signals synchronized with the rotational speed, and in practical engineering applications, the vibration signal of rotating machinery is periodic, so that the unwanted noise signal can be suppressed by TSA. The calculation process is obtained as,

$$y(n) = \frac{1}{N} \sum_{k=0}^{N-1} x(n - kT / \Delta t)$$

where, $x(n)$ denotes the original signal, N is the averaging times, T is the period that selected according to the operating condition of the equipment, and Δt represents the time interval.

The schematic diagram of the TSA is shown in figure 1, there exists lots of noise in the original signal and after the TSA the signal-to-noise ratio of the signal increases because the noise in the original signal is non periodic.

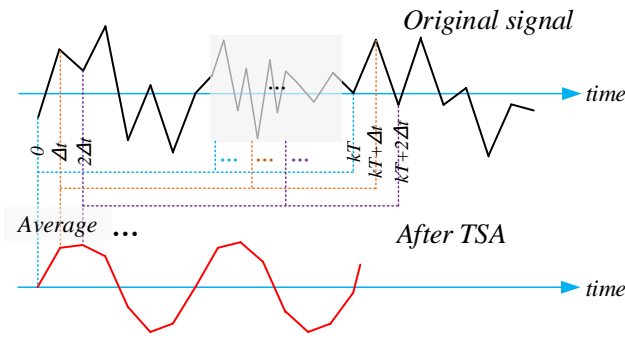


Figure 1: Schematic diagram of TSA

2.2. Spectral entropy (SE)

In order to detect the most fault-related sub-band signal for further analysis, SE method is applied to the spectrum of the vibration signal. A uniform spectrum with broader spectral component (e.g. white noise) yields a high SE value, on the contrary, a narrow spectrum with a few spectral component achieves a low SE value [8]. And it is known that when local gear fault occurs, sidebands with equal interval around meshing frequency or its harmonics will be observed [9]. Therefore SE value can be used to find the most fault-related sub-band in the spectrum. The calculation process of SE of a sub-band in spectrum is given as follows,

$$SE_{sub} = -\frac{1}{\log(L)} \sum_{f=a}^b S(f) \log[S(f)]$$

where L denotes the number of the spectral lines in the sub-band from a to b , and $S(f)$ can be obtained by performing fast Fourier transformation of $x(t)$.

2.3. Envelope analysis (EA)

In case of gear fault occurs, impulsive shocks are modulated by high frequencies, this leads to complex sidebands in frequency spectra, which increases the difficulty of fault detection. EA is one of the most effective demodulation techniques to detect these impulses by removing the carrier signal. The calculation process of EA as follows.

First, Hilbert transformation is performed.

$$\hat{x}(t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(t-\tau)}{\tau} d\tau$$

Then the analytic signal of $x(t)$ is computed.

$$x_a(t) = x(t) + jx(t) = A(t)e^{j\phi(t)}$$

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)}$$

After that the envelope spectrum can be obtained by performing fast Fourier transformation of the envelope signal $A(t)$.

2.4. Fault diagnosis procedure

Figure 2 illustrates the fault diagnosis procedure based on the proposed hybrid approach. First of all, the de-noised signal is obtained by performing TSA technique to the raw signal. Then the spectrum of the de-noised signal is calculated and divided into several sub-band. Secondly, the reciprocal of SE value of each sub-band spectrum is obtained, then an infinite impulse response (IIR) digital filter is designed according to the maximum sub-band boundary, after that the optimal sub-band signal is filtered out. Thirdly, Hilbert transformation is applied to get the envelope spectrum, and fault diagnosis result is derived.

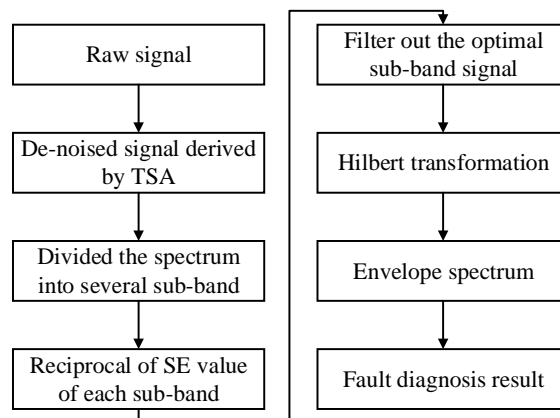


Figure 2: Flow chart of the fault diagnosis procedure based on the proposed hybrid approach

3. EXPERIMENT

3.1. Test rig setup

Figure 3 shows the gearbox test rig setup, which mainly consists of a two-stage gearbox, a motor, a magnetic powder brake, several flexible couplings and control unit. The rotating speed of the motor and the load of magnetic powder brake are controlled by a frequency converter and a load controller respectively, which allow the tested gear to operate under various speeds and loads. The gear ratio of the two-stage gearbox is 3.59, with 23/39 for the first stage, and 25/53 for the second stage. Table 1 gives the design parameters of gears. A small spalling fault is seeded in the second-stage gear pair as shown in figure 3, the depth (d_s), width (w_s) and length (l_s) of the spalling fault are 1mm, 3 mm and 15 mm, respectively.

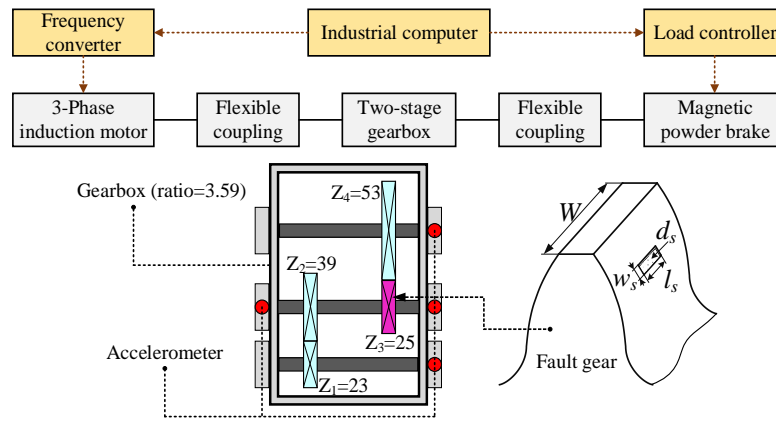


Figure 3. Schematic of the test rig setup

Table 1: The parameters of gears

| Parameter | Value | Parameter | Value |
|------------------|-------|----------------|-------|
| Module | 3 | Pressure angle | 20° |
| Tip clearance | 0.25 | Addendum | 1 |
| Tooth width (mm) | 60 | | |

3.2. Data acquisition

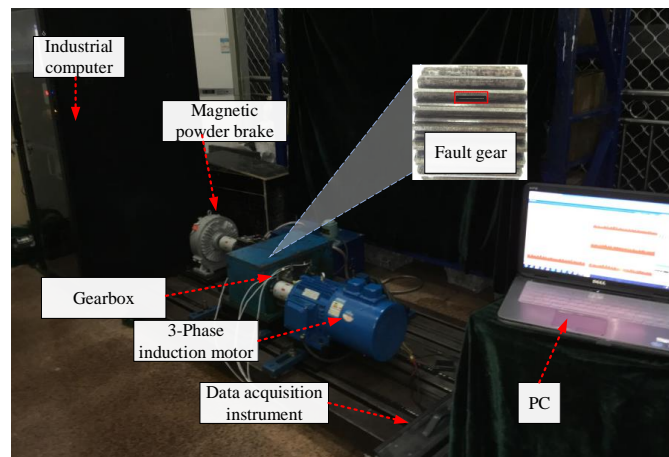


Figure 4: Photograph of the data acquisition system

Four accelerometers are mounted on the bearing casing of the gearbox, a LMS data acquisition system (DAQ) and a PC are used to acquire the vibration data for further processing, figure 4 is the photograph of the data collection process. Vibration data is acquired under 2 rotating speeds and 2 loading conditions for both normal and spalling, as listed in table 2. The signal length for each dataset is 32 seconds. The sampling frequency is 5120 Hz. It is noted vibration signal of the accelerometer mounted on the bearing casing of the second stage is analyzed in this study.

Table 2: Detailed description of various data collection

| Fault severity | Motor speed (rpm) | Load (Nm) | Time length (s) |
|----------------|-------------------|-----------|-----------------|
| Normal | 700/900 | 0/20 | 32 |
| Spalling fault | 700/900 | 0/20 | 32 |

4. RESULT AND DISCUSSION

In order to validate the effectiveness of the proposed hybrid approach, the vibration signal acquired in 700 rpm and 0 Nm working condition is analysed as an example. The time domain waveform of both normal condition and spalling fault condition and their corresponding spectral are shown in figure 5 (a) to (d), it is observed that the spalling fault condition yields higher amplitudes than normal condition in both time domain waveform and spectrum. Some impulsive shocks can be found in the zoomed plot of the waveform of the spalling fault vibration signal as shown in figure 5 (e), and there are abundant sidebands around the meshing frequencies of both first stage gear pair (f_{m1}) and second stage gear pair (f_{m2}), as shown in figure 5 (f). However, it can be also found that some impulsive shocks are fuzzy because they may be contaminated by noise in figure 5 (e). Besides, the sidebands are also fuzzy around f_{m1} and f_{m2} , which may increase the risk of wrong diagnosis results.

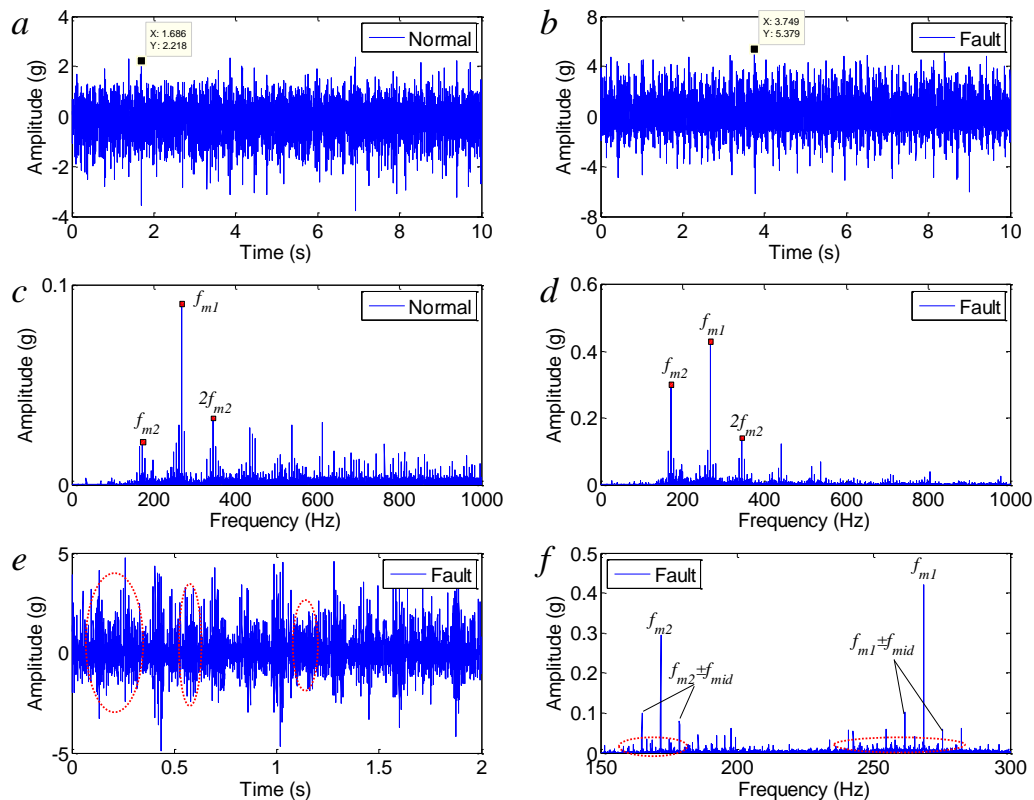


Figure 5: (a) Waveform of normal vibration signal (b) waveform of fault vibration signal (c) spectrum of normal vibration signal (d) spectrum of fault vibration signal (e) zoomed plot of the waveform of fault vibration signal and (f) zoomed plot of the spectrum of fault vibration signal in 700rpm and 0Nm working condition

Figure 6 shows the (a) time domain waveform (b) spectrum and (c) zoomed plot of the spectrum of fault vibration signal after TSA. It is found that all the impulsive shocks become obvious after TSA. And there are also some changes of the structure of the spectrum, the amplitudes of the second meshing frequency and their sidebands become more obvious than in figure 5 (d). It seems that the sidebands around the meshing frequencies become clear in the zoomed plot.

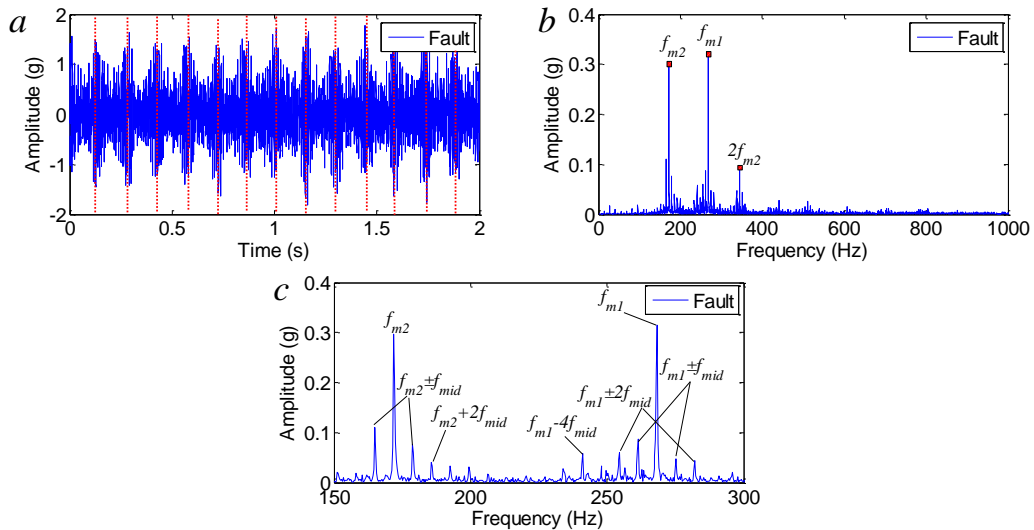


Figure 6: (a) Waveform (b) spectrum and (c) zoomed plot of the spectrum of fault vibration signal after TSA in 700rpm and 0Nm working condition

In order to find the optimized sub-band parameters for EA, the maximum of the reciprocal of SE value is searched in a given grid, and the grid is determined according to the frequency center of $F_1 = [172, 268, 344, 516, 536]$ and $B_1 = [10:2:40]$ for 700 rpm working condition and $F_1 = [221, 345, 442, 662, 690]$ and $B_1 = [10:0.02:0.1]$ for 900 rpm working condition, respectively. Where, F_1 and F_2 are the meshing frequencies of both first stage gear pair and second stage gear pair and their harmonics. Figure 7 denotes the optimal sub-band is (1, 3) which means the corresponding frequency range is $[221-14, 221+14]$.

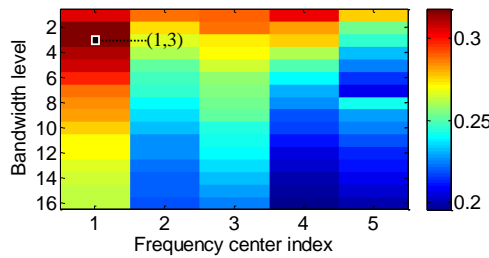


Figure 7: The reciprocal of SE value around meshing frequencies and their harmonics in 700rpm and 0Nm working condition

Figure 8 (a) shows the envelope spectrum of the optimal sub-band vibration signal, figure 8 (b) shows the envelope spectrum of the original vibration signal. It can be observed that only the fault frequency (f_{mid}) can be found in figure 8 (a), however there are lots of fault-unrelated frequencies in figure 8 (b), i.e. rotating frequency of the input shaft (f_i) and output shaft (f_o).

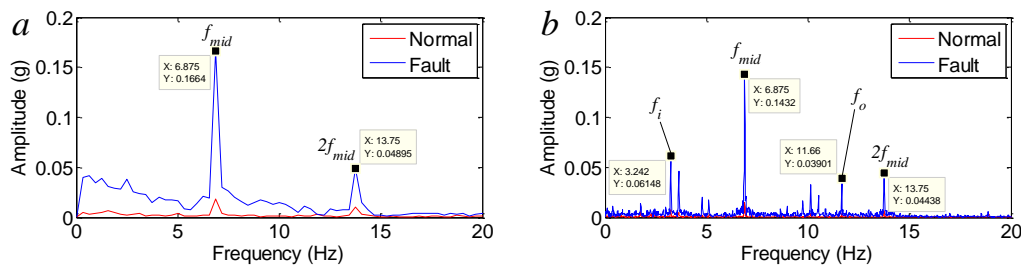


Figure 8: (a) Envelope spectrum of the vibration signal after TSA and (b) envelope spectrum of the original vibration signal in 700rpm and 0Nm working condition

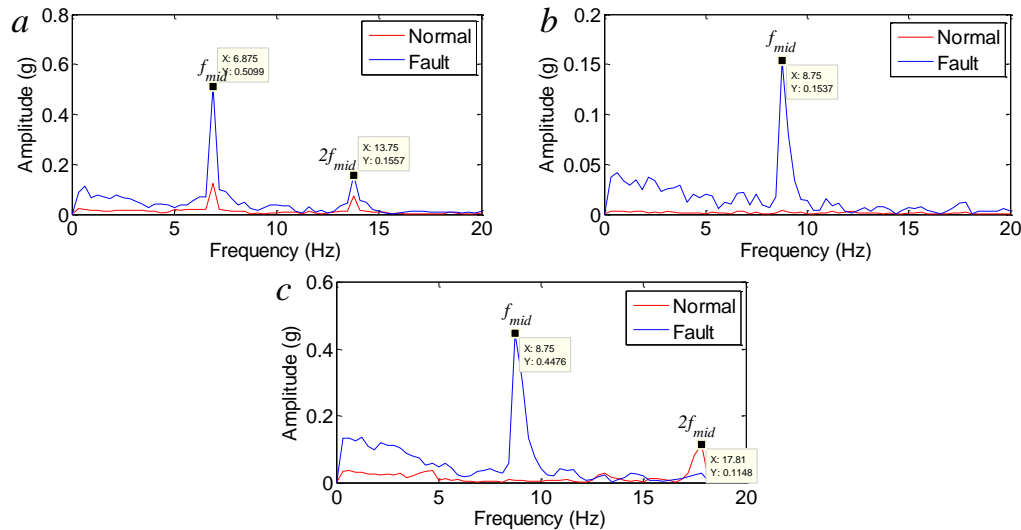


Figure 9: Envelope spectrum of the vibration signal after TSA in (a) 700rpm and 20Nm working condition (b) 900rpm and 0Nm working condition and (c) 900rpm and 20Nm working condition

Figure 9 illustrates the envelope spectral of the optimal sub-band vibration signals in the other working conditions, the fault frequency components are also clear. Therefore, the proposed hybrid approach can help reduce the risk of wrong fault diagnosis results.

5. CONCLUSION

In this paper, a hybrid approach based on time synchronous averaging (TSA), spectral entropy (SE) and envelope analysis (EA) is developed for fault diagnosis of saplling faults. The interference of environmental noise and other components is suppressed by TSA firstly, then the demodulation band for EA is selected based on an optimized SE technique. It can be found that after the pre-processing by TSA, obvious impulsive shocks in time domain waveform as well as clear sidebands in spectrum can be found. Furthermore, spalling fault characteristic frequency components also become clear in the envelope spectrum obtained by SE and EA.

ACKNOWLEDGEMENT

The authors are grateful for the financial support provided by the China North Vehicle Research Institute and the National Natural Science Foundation of China under Contract No.51475053.

REFERENCES

- [1] Lewicki D G. Gear Crack Propagation Path Studies - Guidelines for Ultra - Safe Design. *Journal of the American Helicopter Society*, 2002, 47(1): 64-72.
- [2] Fan X, Zuo M J. Gearbox fault detection using Hilbert and wavelet packet transform. *Mechanical systems and Signal processing*, 2006, 20(4): 966-982.
- [3] Sun H, Zi Y, He Z. Wind turbine fault detection using multiwavelet denoising with the data-driven block threshold. *Applied Acoustics*, 2014, 77: 122-129.
- [4] Tabrizi A, Garibaldi L, Fasana A, et al. Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Meccanica*, 2015, 50(3):865-874.
- [5] Zhao M, Jia X. A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. *Mechanical Systems and Signal Processing*, 2017, 94: 129-147.
- [6] Braun S. The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing*, 2011, 25(4): 1087-1102.

- [7] LaBerge K E, Ehinger R T, Fetty J. Planetary gearbox fault detection using vibration separation techniques. National Aeronautics and Space Administration, Glenn Research Center, 2011.
- [8] Bachiller A, Díez A, Suazo V, et al. Decreased spectral entropy modulation in patients with schizophrenia during a P300 task. *European archives of psychiatry and clinical neuroscience*, 2014, 264(6): 533-543.
- [9] Ma R, Chen Y, Cao Q. Research on dynamics and fault mechanism of spur gear pair with spalling defect. *Journal of Sound and Vibration*, 2012, 331(9): 2097-2109.

Bogie Speed Estimation and Signal Source Separation via Rail Vibration Analysis

G. Qin^{1,3}, F. Gu², Y. Xu², F. Liu² and A. D. Ball²

¹ School of Mechanical Engineering, Hunan International Economics University, Changsha, Hunan, 410205, China

² Centre for Efficiency and Performance Engineering, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

³ Hunan Vtall Information Technology Co., Ltd, Changsha, Hunan, 410100, China

ABSTRACT

As an important part of high speed trains, any fault or failure of the bogie may cause fatal disaster and/or expensive lost. Considering on-train condition monitoring system is still restricted by many limits, bogie condition monitoring via rail vibration may be very valuable. For the rail vibration signals exited by bogies are short termed, non-periodic, non-stationary, not pre-separated corresponding to wheels, and with time-shifting of phases, they must be separated corresponding to each wheel for fault location. In this paper, to separate and reconstruct the vibration signals from two bogie wheels, a measurement mixing model is established and some assumptions about bogie speed, measure accordance, etc. are given. Based on envelope correlation analysis of signals from multi-sensors, the bogie speed and measurement windows are estimated so that the rail vibration signals can be separated corresponding to each bogie wheel, which allows the contact dynamics of the wheel-track to be characterized accurately for the purpose of diagnostics. A simulation and an experiment have been applied to verify this method. Some result as well as discussion will be given.

Keywords: Signal source separation, speed estimation, rail vibration, condition monitoring, bogie

Corresponding author: G. Qin (email: qgj@nudt.edu.cn, qgjnudt@163.com)

1. INTRODUCTION

Bogie, including frame, drive motors, gearbox, axle box, wheels, springs, etc., is a very important part of high speed train. For high-speed locomotive, the electric power from pantograph is firstly feed to an inverter motor and transformed to torque. After the rotating speed reduced by a gearbox, the power is transferred to the axle and wheels.

Reliability in the service of high speed train is very important. Faults or failures in bogie may cause fatal disaster and/or very high expensive lost. Similar with other important equipment, there are three main approaches to insure the reliability of high speed train. The first one is by the high quality of manufacture. But as known in engineering, no matter how low the defect rates of materials and manufacture are, they can never be zero. And the degeneration is sustaining second by second. So the fault and failures can never be eliminated only by reliability design and manufacture. The second approach for reliability assurance is scheduled checks and periodical maintenance. For bogies, their disintegrate detection period is usually 1 or 2 years. Although sometimes a weak fault, such as a small corrosive pitting or brinelling on the surface of a bearing, has been detected, the component may not be changed by a new one for the saving of operating costs. Thus, there will be a potential failure in its running.

The most effective method may be bogie online condition monitoring. In fact, there have been some commercialized CM (Condition Monitoring) systems running on the trains in Europe, Japan and China. For high speed train, although such systems, for example, TDS (Train Detection System), have applied many years, they are still much good at the fault detection of electric or electronic LRU (Line Replaceable Unit).

In recent years, some researches have been done to improve the performance of mechanical CM system for high speed train bogie and track. Most of the researchers have tried to fix the sensors in/on the bogie, built the dynamics model of bogie-track, and analyze the vibration signals to identify the condition of the rail or recognize the faults of bogie. For example, Lee, et al. [1] described a method of estimating irregularities in railway tracks using acceleration data measured from high-speed trains, and discussed the relationship between the mounted location of the accelerometers and the estimated track irregularities. Wei, et al. [2] installed 18 fiber Bragg grating sensors in a train bogie, and used an interrogation system to capture the sensor data. Ortiz, et al. [3] used piezoelectric transducers to transform the vibrations experienced by the bogie into energy that can be used to power the sensors. Trilla and Gratacòs [4] developed and deployed a non-intrusive solution based on a small wireless sensor network that can be installed on the different parts of the bogie and along the whole train, etc.

Although vibration measurement through sensors in bogie is an applicable approach for CM, it is still restricted by many conditions, such as sensor reachability, signal processing, false alarming, data storage and transfer distortion, expenses, etc.

Bogie condition monitoring via rail vibration can take the advantage of easy sensor fixing and data transferring, as well as much cheaper expenses than on-train system, but signal processing will be much difficult because of the weak feature signals and blind source mixture.

Considering the measurement of rail vibration excited by bogie wheels are short termed, non-periodic, non-stationary, not pre-separated corresponding to wheels, and with time-shifting of phases, the basic goal for signal processing, reconstructing the original vibration signals to almost similar with that measured on the bogie in a time interval, is a typical problem of blind moving source separation [5].

In this paper, to separate and reconstruct the vibration signals from two bogie wheels, a measurement mixing model is established and some assumptions about bogie speed, measure accordance, etc. are given. Based on envelope correlation analysis of signals from multi-sensors, the bogie speed and measurement windows are estimated so that the rail vibration signals are separated corresponding to each bogie wheel, which allows the contact dynamics of the wheel-track to be characterized accurately for the purpose of diagnostics. A simulation and an experiment have been applied to verify this method. Some result as well as discussion will be given.

2. MEASUREMENT MODEL AND ITS SOLUTION

2.1. Model of signal measurement

To simplify the problem, only a bogie with 4 wheels and the sensors on one rail are considered in the following part. Assume $s_i(t)$, $i=1,2$, are the original vibration signals from two wheels, $x_j(t)$, $j=1,\dots,M$, are the measured signals, and

$$x_j(t) = s_1(t)w_{1j}(t) + s_2(t)w_{2j}(t) \quad (1)$$

where $w_{ij}(t)$, $i=1,2$, $j=1,\dots,M$, is the measurement window of j th sensor in receiving of $s_i(t)$, M is the amount of sensor.

To reconstruct the signal $s_i(t)$, $i=1,2$, from $x_j(t)$, $j=1,\dots,M$, three reasonable assumptions were supposed as: (a) The speed of bogie could be considered invariant during the measurement. (b) After normalization, all sensors have the same dynamics and the effect of sensor location on the rail can be ignored. (c) The difference of two original signals is small, or there exist some same main components easy to be extracted from measured signals.

Based on these assumptions, measurement windows could be considered having the relationships as Eq. (2) and (3).

$$w_{2j}(t) = w_{1j}(t - \tau) = w_{1j}\left(t - \frac{d_0}{v}\right), \forall j \quad (2)$$

where τ is time interval between the arrival of two wheels to sensor j , d_0 is the axle-distance of two wheels, and v is the speed of bogie.

$$w_{i,j+1}(t) = w_{i,j}(t - \gamma_{j,j+1}) = w_{i,j}\left(t - \frac{d_{j,j+1}}{v}\right) \quad (3)$$

where $\gamma_{j,j+1}$ is time delay between sensor j and sensor $j+1$; $d_{j,j+1}$ is the distance of these two sensors.

Based on the assumptions, the mixing model of signal measurement can be re-written as

$$\begin{aligned} x_1(t) &= s_1(t)w(t - t_0) + s_2(t)w(t - t_0 - \frac{d_0}{v}) \\ x_2(t) &= s_1(t)w(t - t_0 - \frac{d_{12}}{v}) + s_2(t)w(t - t_0 - \frac{d_0}{v} - \frac{d_{12}}{v}) \\ &\vdots \\ x_M(t) &= s_1(t)w(t - t_0 - \frac{d_{1M}}{v}) + s_2(t)w(t - t_0 - \frac{d_0}{v} - \frac{d_{1M}}{v}) \end{aligned} \quad (4)$$

where t_0 is the center of measurement window for $s_1(t)$. d_{1j} , $j=1 \sim M$, is the distance between sensor 1 and sensor M .

Solving this problem is equals to finding:

$$\begin{cases} s_{1j}(t), t \in \left[t_0 + \frac{d_{1j}}{v} - \frac{L}{2}, t_0 + \frac{d_{1j}}{v} + \frac{L}{2} \right], j = 1, 2, \dots, M \\ s_{2j}(t), t \in \left[t_0 + \frac{d_0 + d_{1j}}{v} - \frac{L}{2}, t_0 + \frac{d_0 + d_{1j}}{v} + \frac{L}{2} \right], j = 1, 2, \dots, M \\ t_0, v, L \end{cases} \quad (5)$$

where L is the width of measurement window for signal $s(t)$.

2.2. Solution procedure

To reconstruct the signals $s_i(t)$, $i=1,2$, extract the feature and detect the fault in bogie, the measurement window function $w(\cdot)$, parameter τ , γ or v must be estimated, and some effective reconstruction algorithm should be applied.

The procedure for signal reconstruction and fault detection is illustrated in figure 1, in which the form of measurement window functions is estimated firstly. Then, based on Eq. (4), the bogie speed can be estimated and the signals separated corresponding to wheels at the same time. Finally, the original vibration signals can be reconstructed and the fault be detected through the conventional approaches.

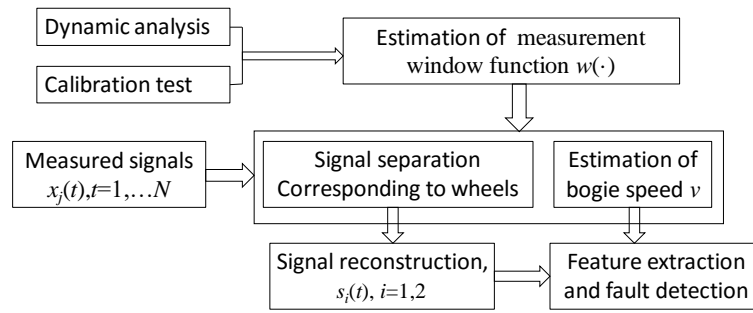


Figure 1. Solution procedure

3. ESTIMATION OF TEST WINDOWS

In the estimation of measurement window, both its form and parameter should be determined or identified. Although the best approach may be combining dynamic analysis with calibration test (including system identification), in the preliminary study, the function's form is just predetermined by the shape of signals measured in calibration tests.

3.1. Form of measurement window

From the shape of signals measured in calibration test by rolling a wheel on the rail, a window's form of Gaussian function is written as

$$w(t) = e^{-\frac{(t-t_0)^2}{\alpha^2}} \quad (6)$$

where $\alpha > 0$ is the width parameter, t_0 is the center of window.

3.2. Identification of measurement window

To identify the parameter α and t_0 in Eq. (6) and Eq.7, the estimation procedure is illustrated in figure 2. The RMS (Root Mean Square) -based envelope of vibration signal is calculated firstly. Through the correlation analysis between the 4 power of envelopes from different accelerometer as well as that between each envelope and the window function, the bogie speed and the parameter α , t_0 can be estimated, respectively.

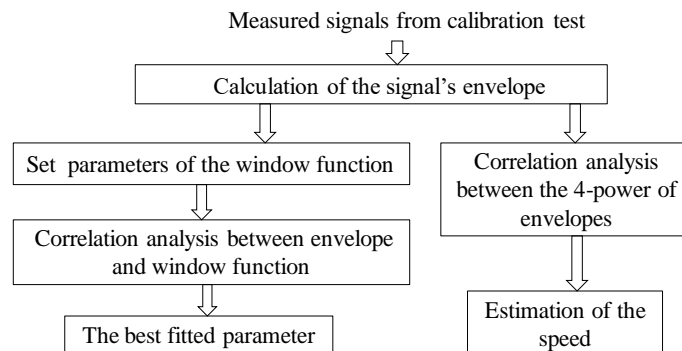


Figure 2. Procedure of measurement window identification

3.3 Calibration Tests

There are two easy-realized test methods to calibrate the time window, rolling a wheel on the rail or knock the rail by an impulse hammer near the sensors. Here a pair wheels was used to roll through a section of straight rails, under which three accelerometers were fixed, as shown in figure 3(a)~(c).

The rail vibration signals from three accelerometer and their envelopes when the wheels rolling forward and backward are illustrated in figure 4. The best correlation coefficient between measurement window with each envelop were calculated between the test intervals of 0-5s, 5-7.8s, 7.8-10.6s, 10.6s-13.6s, and 13.7-16.7s. As an example, for the 3 signals between 0-5s, the best correlation coefficients are 0.9906, 0.9925 and 0.9934, while their corresponding window width parameter α are all 1.6667, which means the results are unanimous. The speed estimations for 5 tests are also illustrated in table 1, where the values in the row marked with “pre-fit” are calculated just by the correlation analysis of vibration signal envelopes and that in the row with “post-fit” are estimated by the correlation analysis of identified windows.

From the results, it can be seen that: (a) Gaussian type measurement window can be applied to calculate the bogie's speed, and the directions of bogie's movement are all estimated correctly. (b) The speed estimation depends on the width applied in envelope calculation. The criteria for best width selection should be considered. (c) A dynamic model of rail vibration may do good to determine of the window's form.

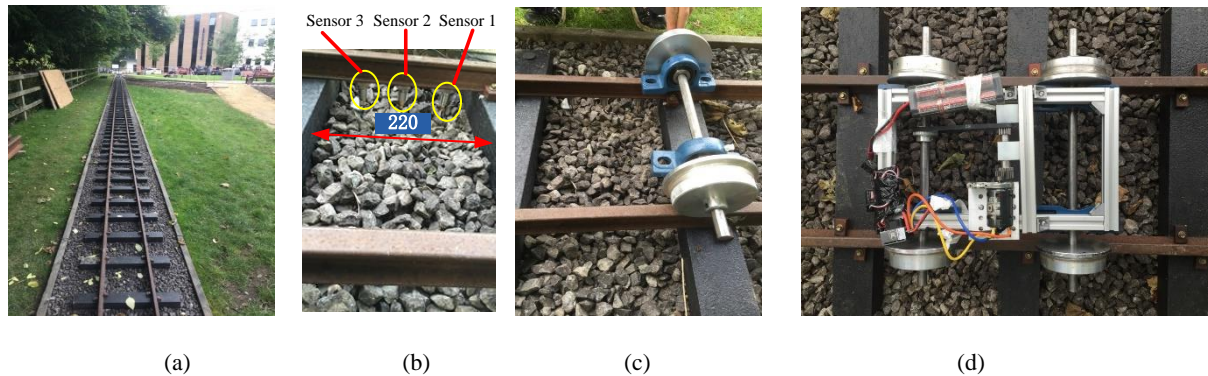


Figure 3. Straight rails, accelerometers, a pair of wheels rolling on the rail and experiment bogie

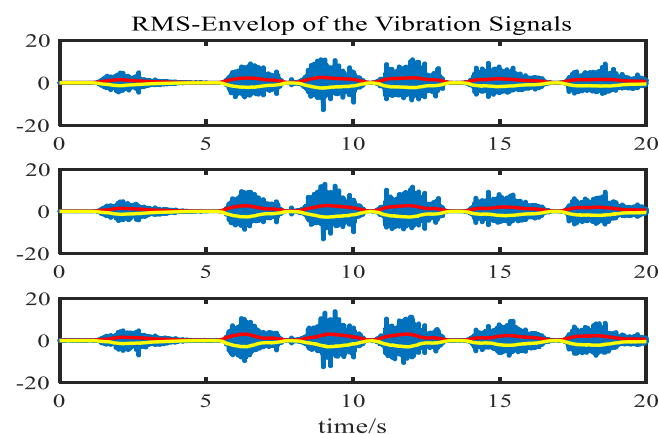


Figure 4. Rail vibration signals & their envelopes when the wheels rolling forward and backward

Table 1: Speed estimation of 5 tests (m/s)

| | 0-5s | 5-7.8s | 7.8-10.6s | 10.6-13.6s | 13.7-16.7s |
|----------|--------|---------|-----------|------------|------------|
| Pre-fit | 0.4209 | -0.9091 | 0.8832 | -1.2968 | 0.7056 |
| Post-fit | 0.4903 | -0.6617 | 0.8081 | -0.8949 | 0.4746 |

4. SIGNAL SEPARATION AND BOGIE SPEED ESTIMATION

4.1. Algorithm for signal separation

The algorithm for bogie speed estimation and signal separation via rail vibration can be listed as follows:

Step 1: Calculate the envelope $\hat{E}_j(t)$ of $x_j(t)$, $j=1 \sim M$

Step 2: Estimate the bogie's speed v through correlation between $\left[\hat{E}_j(t)\right]^4$.

Step 3: Set a solution searching domain, and look for the best estimation for t_0 and α by

$$\hat{E}_j(t) = w(t - t_0 - \frac{d_{1j}}{v}) + w(t - t_0 - \frac{d_0}{v} - \frac{d_{1j}}{v}), j=1, \dots, M$$

Step 4: Set all the windows $w(t - t_0 - \frac{d_{1j}}{v})$ and $w(t - t_0 - \frac{d_0}{v} - \frac{d_{1j}}{v})$, $j=1, \dots, M$

Step 5: Calculation of $s_{ij}(t)$, $i=1,2$; $j=1, \dots, M$, where

$$s_{1j}(t) = \frac{w(t - t_0 - \frac{d_{1j}}{v})x_j(t)}{w(t - t_0 - \frac{d_{1j}}{v}) + w(t - t_0 - \frac{d_0}{v} - \frac{d_{1j}}{v})}, \quad t \in \left[t_0 + \frac{d_{1j}}{v} - \frac{L}{2}, t_0 + \frac{d_{1j}}{v} + \frac{L}{2} \right]$$

$$\text{and } s_{2j}(t) = \frac{w(t - t_0 - \frac{d_0}{v} - \frac{d_{1j}}{v})x_j(t)}{w(t - t_0 - \frac{d_{1j}}{v}) + w(t - t_0 - \frac{d_0}{v} - \frac{d_{1j}}{v})}, \quad t \in \left[t_0 + \frac{d_0 + d_{1j}}{v} - \frac{L}{2}, t_0 + \frac{d_0 + d_{1j}}{v} + \frac{L}{2} \right].$$

For Gaussian window, $L=4\alpha$.

4.2. Simulation of vibration signals with bearing fault

To verify this algorithm, a time series of vibration signal with bearing fault in one wheel was simulated as

$$x_1(t) = s_1(t)w(t-t_0) + s_2(t)w\left(t-t_0 - \frac{d_0}{v}\right)$$

$$x_2(t) = s_1(t)w\left(t-t_0 - \frac{d_{12}}{v}\right) + s_2(t)w\left(t-t_0 - \frac{d_0}{v} - \frac{d_{12}}{v}\right)$$

$$x_3(t) = s_1(t)w\left(t-t_0 - \frac{d_{13}}{v}\right) + s_2(t)w\left(t-t_0 - \frac{d_0}{v} - \frac{d_{13}}{v}\right)$$

$$\text{where } s_1(t) = \sum_{i=0}^M \left(A_0 \cos(2\pi f_m(iT + \tau_i)) \right) \left(e^{-B(t-iT-\tau_i)} \cos(2\pi f_n(t-iT-\tau_i)) \right) + n(t),$$

$$s_2(t) = n(t) \square N(0,1), w(t) = e^{-\frac{t^2}{\alpha^2}}, d_0=1.2\text{m}; d_{12}=d_{23}=0.13\text{m}; v=0.6\text{m/s}; \alpha=1.$$

The original signal and simulated measurements by 3 sensors were illustrated in figure 5, respectively. After estimation, the window width parameter $\alpha=0.4952$. The separation of simulated signals was shown in Figure 6. Obviously, the separation results are very similar with the windowed original signals.

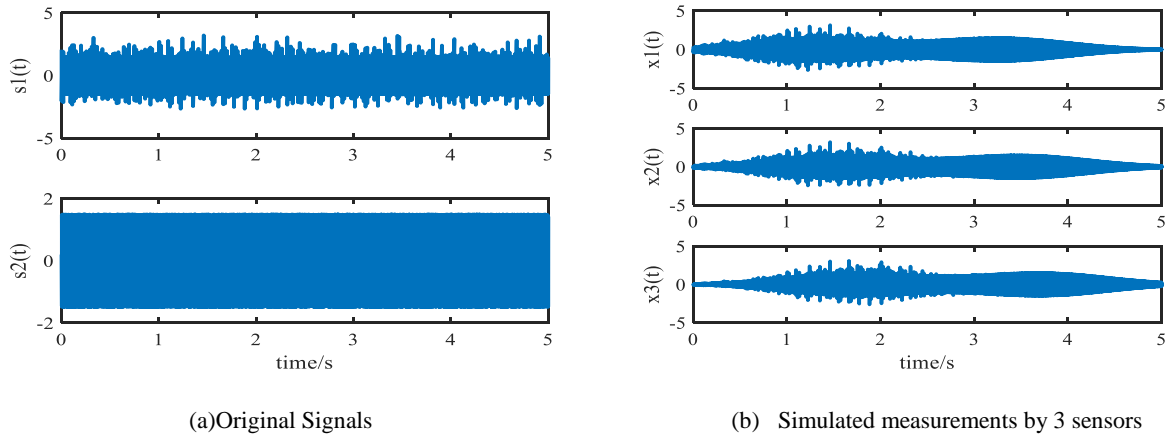


Figure 5. Measurement simulation

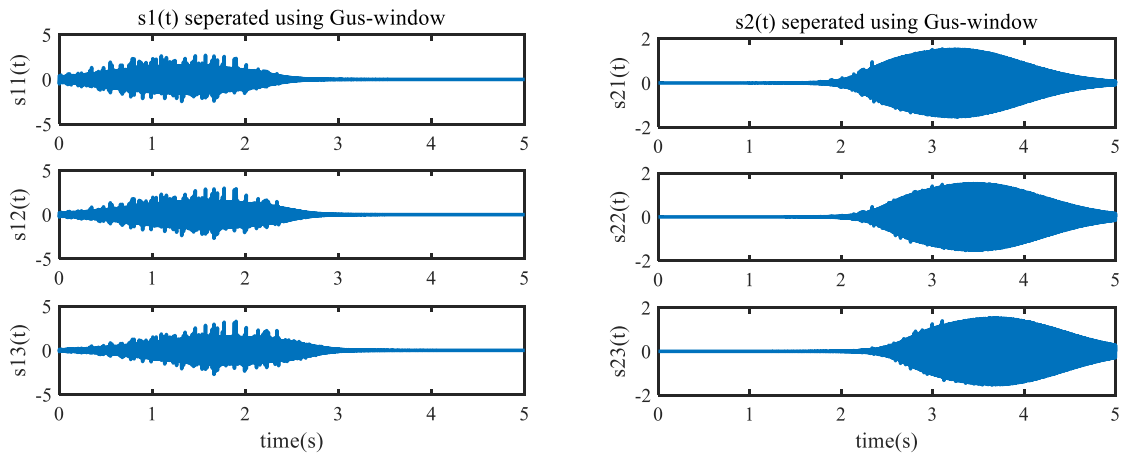


Figure 6. Separation of simulated signals based on Gaussian-windows

4.3 Experiments

When an experiment bogie (as figure 3(d)) running on the rail, the vibration signals recorded by 3 accelerometer are illustrated in figure 7. Based on the estimation of the bogie speed $v=0.82\text{m/s}$ with

window width parameter $\alpha=1.5s$, the vibration signal were separated as figure 8. The two estimated original signal is similar, but with a time delay.

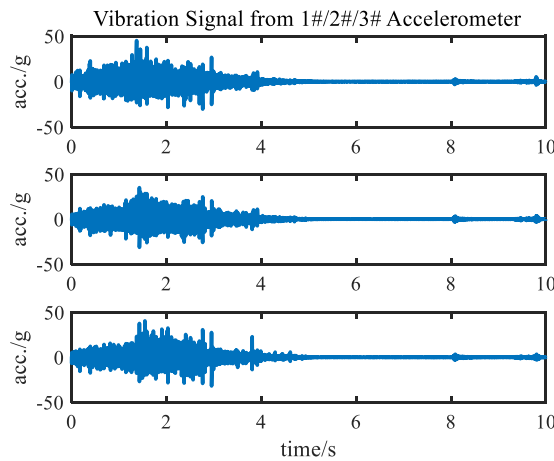


Figure 7. Measured signals.

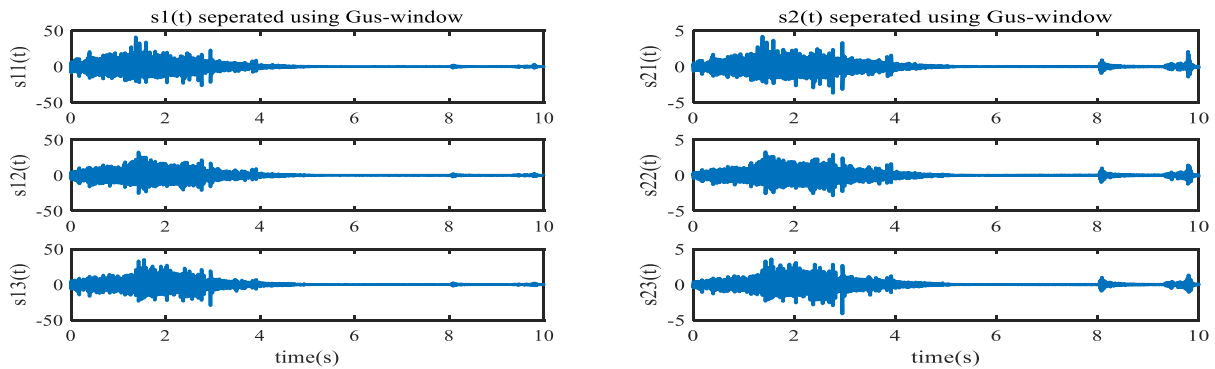


Figure 8. Separation of experiment signals based on Gaussian-windows

5. CONCLUSIONS AND DISCUSSION

For the signals measured from rail vibration that excited by bogie wheels are short termed, non-periodic, non-stationary, not pre-separated corresponding to wheels, and with time-shifting of phases, their separation is a typical problem of blind moving source separation. Based on some assumptions of the bogie speed and measure process, a Gaussian measurement window is selected for the signal mixing model. Through the envelop calculation and correlation analysis, bogie speed and window parameters are both calculated firstly. The vibration signals were separated corresponding to different wheels. From the signal processing and result, it can be seen that: (a) The directions of bogie's movement are all estimated correctly. (b) Accuracy of speed estimation is very important for signal separation. Envelope correlation analysis is much better for speed estimation that the original signal correlation analysis, but it is still not accurate enough, and the result is much affected by signal pre-process and envelops calculation. (c) The accuracy of separation depends the distance of two wheels or the degree of signal mixture. (d)The difficulty of signal separation should be further estimated by signals measured from railway, not only by test rig.

ACKNOWLEDGEMENT

The authors gratefully acknowledge financial support of Chinese NQI Common Technology Research and Application Project (Grant No. 2016YFF0203400), National Natural Science Foundation of China (Grant No. 51375484), Hunan strategic emerging industries major core technology project (Grant No. 2015GK1020) and Hunan Education Research Program (Grant No.14A083).

REFERENCES

- [1] Jun Seok Lee, Sunghoon Choi, Sang-Soo Kim, et al. (2012). A Mixed Filtering Approach for Track Condition Monitoring Using Accelerometers on the Axle Box and Bogie [J]. *IEEE Transactions on Instrumentation and Measurement*, 61(3): 749-758
- [2] Chuliang Wei, Zemin Cai, Hwa-yaw Tam, S.L. Ho, and Qin Xin. (2012). Real-time Fault Diagnosis of Train Bogie using FBG Sensors [P]. *Proc. 2012 International Conference on Intelligent System Design and Engineering Application*: 1087-1090
- [3] [3] Javier Ortiz, Pedro M Monje, Nerea Zabala, et al. (2014). New proposal for bogie-mounted sensors using energy harvesting and wireless communications [P]. *Proc IMechE Part F: J Rail and Rapid Transit*, 228(7): 807–820
- [4] Alexandre Trilla and Pau Gratacòs. (2016). Maintenance of bogie components through vibration inspection with intelligent wireless sensors: A case study on axle-boxes and wheel-sets using the empirical mode decomposition technique [P]. *Proc IMechE Part F: J Rail and Rapid Transit*, 230(5): 1408–1414
- [5] Xiangdong Huang, Xukang Jin, and Haipeng Fu. (2016) Short-Sampled Blind Source Separation of Rotating Machinery Signals Based on Spectrum Correction. *Shock and Vibration*, Volume 2016, Article ID 9564938, 10 pages. <http://dx.doi.org/10.1155/2016/9564938>

Session 7

End User Applications and Maintenance in Industry 1

Session Chair Prof. Andrew Ball

Wind Turbine Bearing Fault Detected with IAS Combined with Harmonic Product Spectrum

(A. Hugo, K. Ilyes, L. Quentin)

A Comparison of Architectures and Evaluation Metrics for Streaming Machine Learning Algorithms in Industry 4.0 Applications *(C. Berry, G. Hall, B. Matuszewski, L. Shark)*

Measurement Uncertainty When Thermal Effects on a Machine Tool Using Non-Contact Displacement Transducers *(S. Gorton, S. Fletcher, D. Clough, A. P. Longstaff)*

Research on Spectrum Components of Planetary Gearbox Vibration Signals by Dynamic Modelling and Filtering *(L. Zhang, N. Hu, F. Gu, Z. Cheng)*

Wind Turbine Bearing Fault Detected with IAS Combined with Harmonic Product Spectrum

André Hugo¹ *, Khelf Ilyes^{2,4}, and Leclère Quentin³

¹ Univ Lyon, UJM Saint Etienne, LASPI, EA 3059, F-42334, IUT de Roanne, France

² Univ Lyon, CNRS INSA-Lyon, LaMCoS UMR5259, F-69621, Villeurbanne,

³ Univ Lyon, INSA-Lyon, LVA EA677, F-69621, Villeurbanne, France

⁴ ENGIE GREEN, 2 place Samuel de Champlain, 92400 COURBEVOIE

ABSTRACT

A few years ago, Instantaneous Angular Speed (IAS) signal analysis has been proven able to detect natural bearing faults. Moreover, it has been applied on a 2MW wind turbine shaft line and proven able to detect low speed shaft unbalance using an encoder located on the high speed shaft turbine generator. This paper will show that a generator bearing fault can also be monitored using an encoder located on the low speed shaft. This step forward is made difficult by the lack of energy in the bearing fault speed fluctuation amongst the multitude of noisy phenomena, an alternative technique is proposed in this paper to extract dry impacts components from IAS spectrum. Harmonic Product Spectrum has originally been proposed in the sixties to detect fundamental frequency in a noisy signal. Still in use in the domain of speech signal processing, this method eases a precise voice pitch tracking. This paper proposes a revision of HPS to IAS analysis (and more generally to vibration monitoring), and a peculiar adaptation to look for inner ring modulated defects. The efficiency of the technique will be finally shown on real measurements issued from a 2MW wind turbine generator bearing fault.

Keywords: Instantaneous Angular Speed, Harmonic Product Spectrum, Bearing Fault, Rotating Machine Monitoring.

Corresponding author: Hugo ANDRE (email: hugo.andre@univ-st-etienne.fr)

1. INTRODUCTION

A few years ago, Instantaneous Angular Speed (IAS) signal analysis has been proven able to detect natural bearing faults [1]. This major experimental demonstration shows that mechanical faults can be detected through the reading of the torsional shaft movement rather than the transverse vibration of the equipment housing. Amongst the different technics that can be used to get IAS signal, Elapse Time seems to bring the best results and is under consideration in this paper. This acquisition method introduces specific limitations which have recently been detailed [2] and which mainly account for the difficulty to use classical vibration signal processing tools. For instance, within many methods that have been developed to improve bearing fault detection, the most favoured one is probably the envelope spectrum analysis of the vibration signal filtered on a conveniently chosen frequency band. However, the adaptation of this technique to Instantaneous Angular Speed (IAS) signals is not straightforward, since neither the impact nor the structural response reach the quantification threshold obtained with current acquisition systems.

This paper proposes an alternative technique to extract dry impacts components from IAS spectrum.

Originally proposed in 1968 by M.R Schroeder [3] and soon after democratized by Noll [4], Harmonic Product Spectrum is a method derived from Cepstrum analysis and used to detect fundamental frequency in a noisy signal. In the domain of speech signal processing, this methods eases a precise voice pitch tracking, which fundamental frequency can be mixed with the noise if not pointed out by its higher harmonics. This paper proposes a revision of HPS to IAS analysis (based on elapse time acquisition technique), and a peculiar adaptation to look for inner ring modulated defects. The efficiency of the technique will be finally shown on real measurements issued from a wind turbine generator bearing fault.

2. PROCESSING TOOLS

2.1. Description of HPS

Harmonic Product Spectrum is a tool dedicated to reveal the fundamental frequency of a harmonic set mixed with noise. The intuitive reasoning for the method is that the peach peaks in the log spectrum add coherently while the other portions of the log spectrum are uncorrelated and add non coherently. The frequency compression results in a sharper final peak as depicted in figure 1. This figure, along with section, is largely inspired from the original paper which proposed HPS for the first time [4]. The antilog version of this schematization is the Harmonic Product Spectrum, and is simply defined such as:

$$\pi(\omega) = \prod_{k=1}^K X(k\omega)$$

With K the number of harmonics taken into account and $X(\omega)$ is the amplitude spectrum of the time signal. Therefore, the HPS is a function of the frequency and its unit is the original spectrum unit power K .

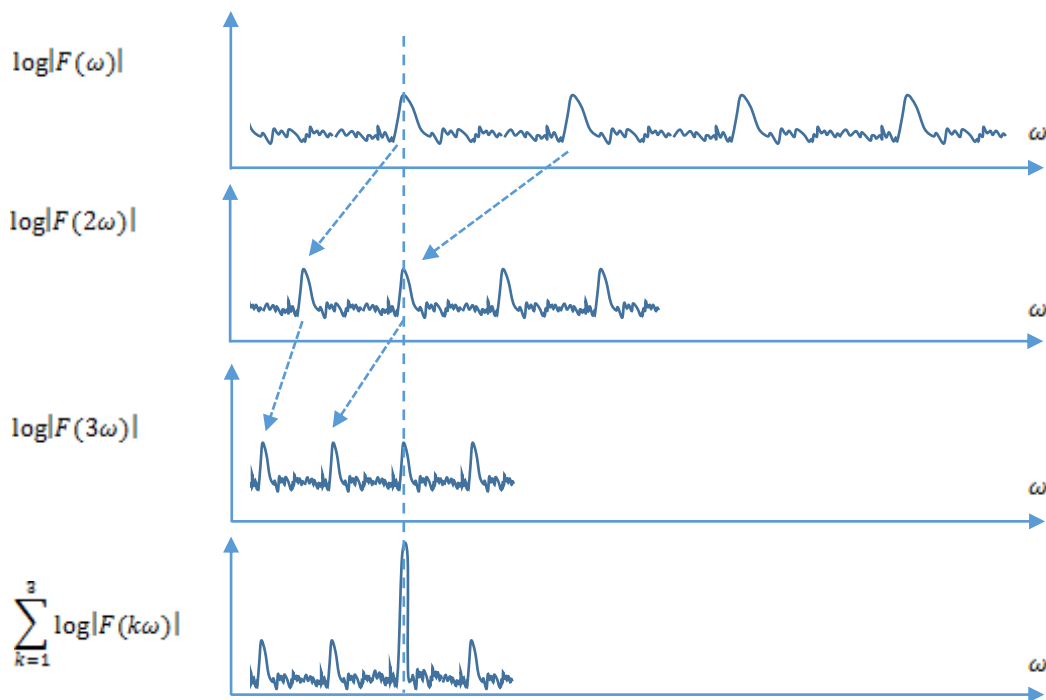


Figure 1 Developpement of the Harmonic Product Spectrum as the antilogarithm of the sum of harmonically compressed log spectra.

2.2. Application to rotating machinery

The main difference between speech analysis and rotating machinery monitoring is that the user is not only looking to locate the harmonic set frequency, but also to apprehend its amplitude in order to estimate a fault severity. Regarding this peculiarity, the fact that HPS unit is the original spectrum unit power K might pose a problem. A first proposal to tackle this issue is to present the HPS as a probability function, such as:

$$\pi(\omega) = \prod_{k=1}^K \frac{X(k\omega)}{\int_0^{f_{max}/K} X(k\omega) d\omega}$$

This slight modification transform the original HPS in a probability function without unit that can be interpreted as the chance for a frequency to correspond to the fundamental frequency of the harmonic set. Still, several difficulties are to be tackled:

1. In case of bearing monitoring, the healthy mode is not supposed to induce a harmonic set in the signal. This will be discussed in the end of the paper.

2. Rotating machinery signals are never populated with only one harmonic set. The more harmonic sets in the signal, the lower the probability will be. It appears therefore necessary to limit the frequency span of interest to a frequency band $[f_1; f_2]$ where no other harmonic set is expected:

$$\pi(\omega) = \prod_{k=1}^K \frac{X(k\omega)}{\int_{f_1}^{f_2} X(k\omega) d\omega}$$

3. Several defect modes are not only a succession of harmonics, they are also modulated by a kinematic frequency which often corresponds to the shaft frequency carrying the defect. This will be solved in the next section.

2.3. Modulated Harmonic Spectrum

The formulation proposed by Noll could be slightly modified to profit from lateral bands induced by modulations such as shaft modulation or encoder geometric faults. For example, inner bearing fault are expected to be shaft modulated since the speed fluctuation they induce on the shaft depend on the load transmitted by the rolling element passing through the defect. Although this was already explained by [6] concerning vibration monitoring with a simple phenomenological model, academics are currently working on justifying a similar behaviour with IAS using a proper and brand new mechanical model [7].

In this work, modulation effect are considered by adding a preliminary step where the original spectrum is multiplied by shifted spectrum similarly to the HPS reasoning, modulated peaks will add coherently on the carrying frequency. The completed development is presented in figure 2 in its log version, in the event that only one side band and three harmonics are considered. The authors propose to call its antilog version the Modulated Harmonic Spectrum (MHPS) and to define it such as:

$$\pi(\omega) = \prod_{k=1}^K \prod_{i=-M}^M |X(k\omega + i\omega_c)|$$

Where M is the number of side bands considered, and ω_c is the modulating frequency. The latter corresponds to the carrying shaft frequency in the case of inner ring fault.

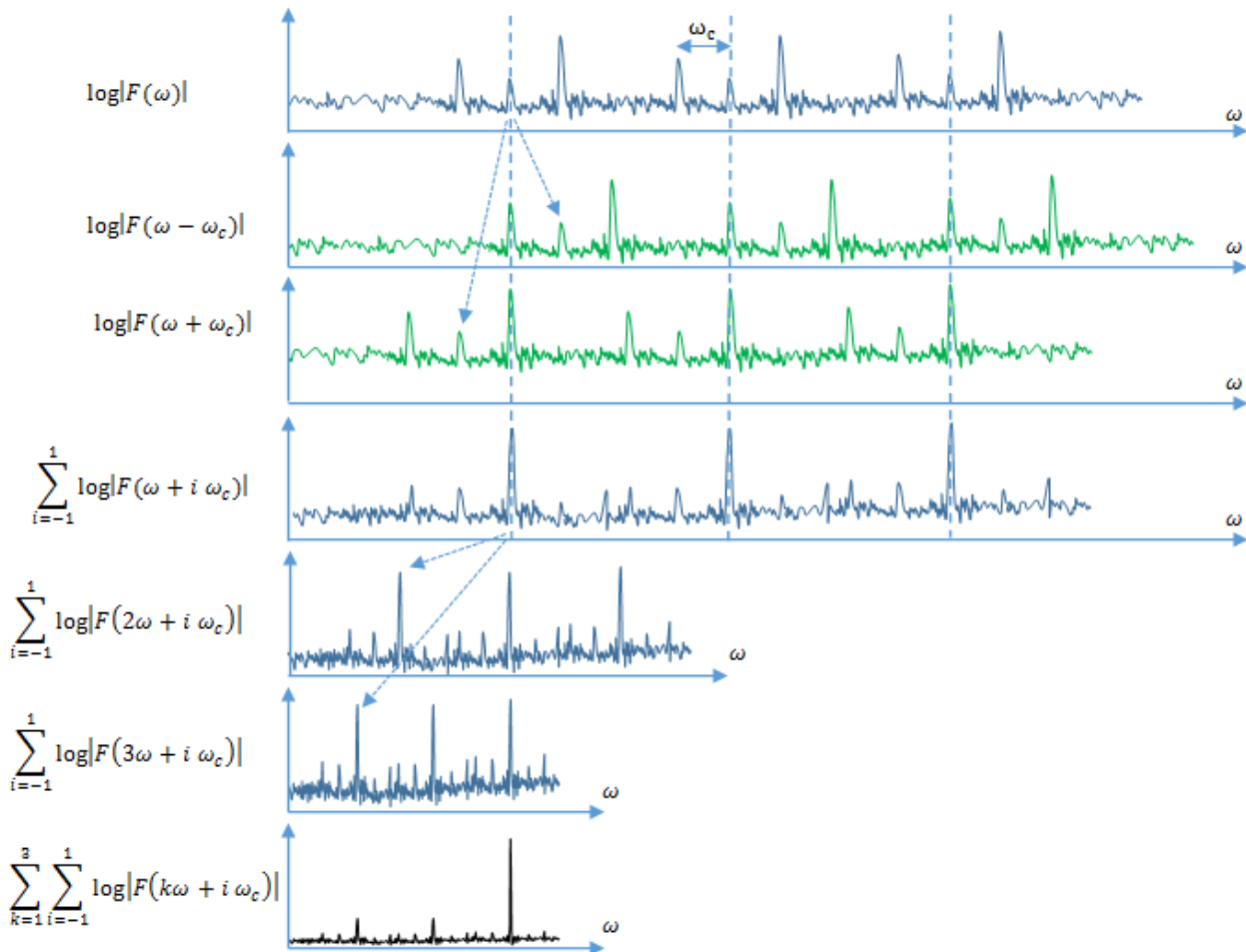


Figure 2 Development of the Modulated Harmonic Product Spectrum as the antilogarithm of the sum of harmonically compressed log spectra which has been demodulated.

3. MATERIAL AND METHODS

3.1. Instantaneous angular speed

Instantaneous Angular Speed (IAS) has recently appeared as an original and promising tool to monitor mechanical parts of rotating machines. Mechanisms running under non stationary conditions, such as wind turbine, are especially suited for this method since the issued signal is intrinsically sampled in the angular domain. Readers interested by the acquisition method can refer to [2]. Measurements have been obtained with the Elapse Time method, on a 20480 pulses per revolution magnetic encoder sampled with a 120MHz counter clock. The aim of this paper is not to present IAS monitoring, especially since HPS might bring interesting results on classical vibration measurements. However, the fact that Elapse Time technique yields an angular sampled signal is important since it helps harmonics components of cyclo-stationary phenomenon to be concentrated in one frequency channel.

3.2. Wind turbine bearing defect

A long term study is being carried over a MM82 wind turbine to qualify the ability of this monitoring system. The wind turbine set-up is presented in figure 3 for the reader to realize the easiness of the involved instrumentation in regards with the kinematic complexity of the turbine line shafting. The IAS signal can be computed from the generator optical encoder, which is a high quality incremental encoder used by the converter to correctly synchronize the asynchronous generator; but also from a lower quality magnetic encoder, installed in retrofit operation on the low speed shaft directly carrying the rotor hub. It has been decided to equip the low speed shaft since the most expensive shaft line elements are cinematically and

physically closer from it. The generator encoder has already been shown able to estimate the amplitude of speed fluctuation induced by a low speed shaft unbalance [5], and as a collateral benefit, this paper will check that the low speed shaft encoder is able to estimate speed fluctuation issued from a generator bearing. Originally, the acquisition card, FPGA type, is embedded in an industrial PC directly installed in the nacelle. The encoder is settled at the rear of the generator, coupled itself to the speed increaser gearing through a drive shaft coupling. The speed transmission is made of one star epicyclical gear train and one parallel stage mounted in serial configuration to obtain a global speed ratio approximately equal to 105.

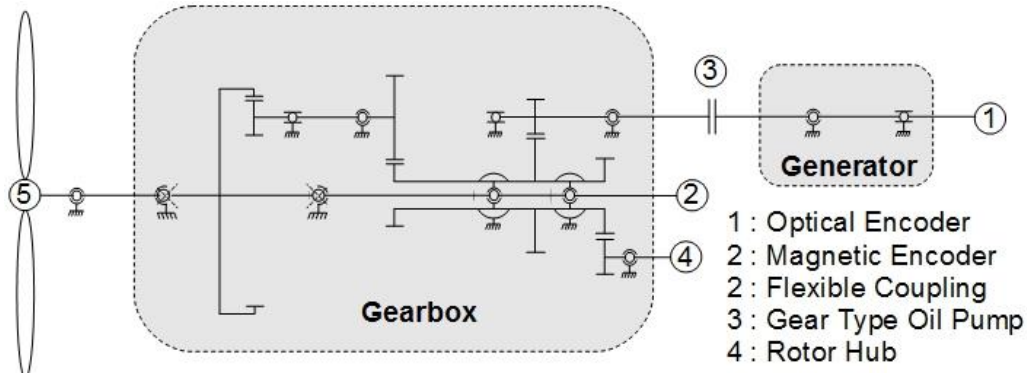


Figure 3: Kinematic Scheme of the wind turbine set-up

One of the two deep groove ball bearings in the generator experienced a double fault during the measurement campaign. The same bearing experienced, in the meantime, an inner ring and an outer ring defect: small spall (1cm²) on the outer ring and even smaller (20mm²) on the inner ring. Cyclic frequency of interest are successively defined here under in terms of events per high speed shaft revolution and events per low speed shaft revolution. Once the signal is sampled in the angular domain, bearing characteristic frequencies do not depend on the speed anymore but only on the reference shaft, ie. The shaft where the encoder is installed.

$$\begin{aligned} f_{BPFI} &= 4.81 (hss)^{-1} = 507 (lss)^{-1} \\ f_{BPFI} &= 3.19 (hss)^{-1} = 336 (lss)^{-1} \\ f_c &= 1 (hss)^{-1} = 105,5 (lss)^{-1} \end{aligned}$$

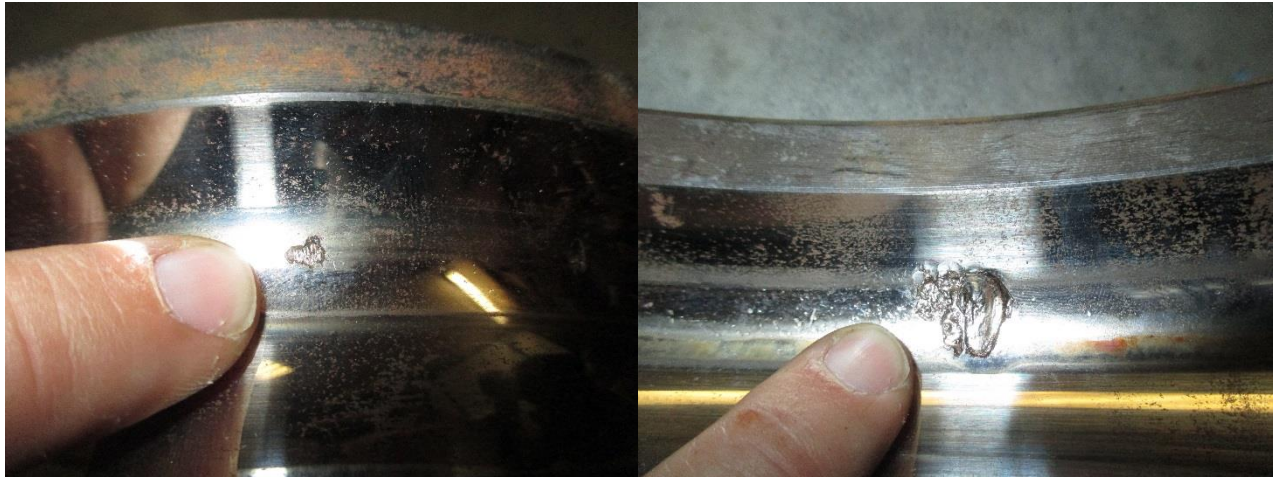


Figure 4: inner ring defect (on the left) and outer ring defect (on the right) of the generator front bearing.

4. RESULTS

4.1. State of the art results

Those defects were both efficiently detected using the optical encoder installed at the back of the generator, leading to an early stage replacement: it was not necessary for the operator to limit the maximal

power of the turbine generator until the spare bearing replacement. The amplitude spectrum obtained with this sensor shew strong emergence on both BPFO et BPFI of the corresponding bearing. The processing techniques used to detect the faults using this very close optical encoder do not hide any peculiar difficulties, and can be found in [5]. This allow us to precisely estimate the actual characteristic frequency of the defect, and take this information into account in the rest of the analysis. This difficulty will be mentioned in the discussion section.

For the sake of conciseness, this paper will focus on the IAS obtained from the low speed shaft magnetic encoder which is 5 meters, 3 gear trains and one mechanical coupling away from the defective generator bearing. It is important to precise that the fuse coupling between the generator and the gearbox is soft in flexion but rigid in torsion. However, as it can be seen on the low speed shaft IAS spectrum focused on bearing characteristic frequencies do not clearly exhibit the defect. BPFI, in particular, will reveal itself if we take focus the same spectrum around some of its harmonics and high speed shaft modulated combs.

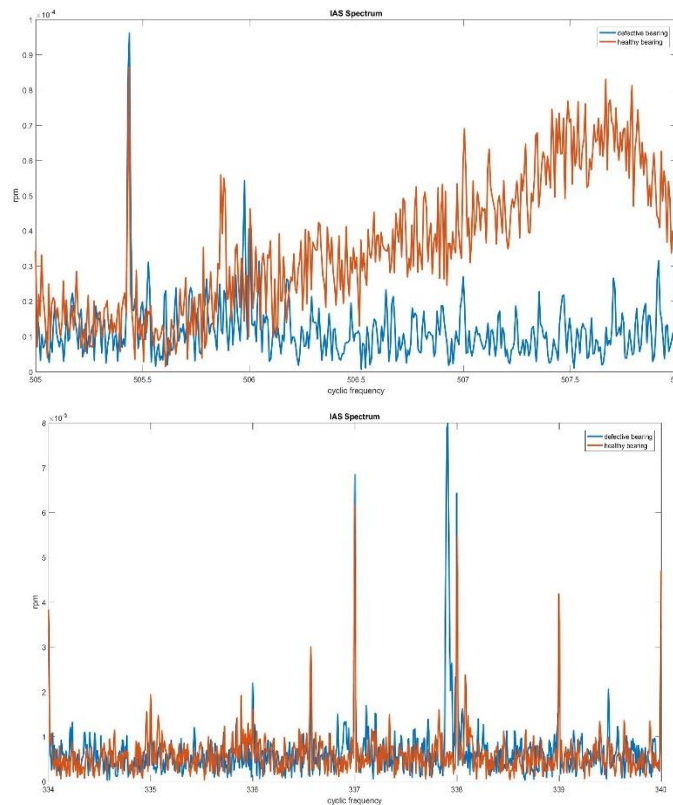


Figure 5: IAS Spectrum focused around BPFI (characteristic cyclic frequency) on left, and on BPFO on right. Red spectrum is healthy bearing, when blue is defective. Cyclic frequency span

4.2. Harmonic Product Spectrum

Harmonic Product Spectrum takes into account harmonics only. Therefore, its ability to detect bearing fault is expected to work better with outer ring faults than inner ring ones. Figure 6 simultaneously shows HPS for defective and healthy signals. 4 harmonics are taken into account ($K = 4$) and the frequency band of interest is $2(lss)^{-1}$ wide: $[f_1; f_2] = [337; 339]$. This means the ordinate axis stands for the probability of each cyclic frequency channel in the interval $[f_1; f_2]$ to host the fundamental frequency of a harmonic set. As a result, a middle scoring peak ($< 40\%$) is seen for $f = 337.89 (lss)^{-1}$ on the defective signal, while a small peak emerges on the healthy signal for $f \sim 338.1 (lss)^{-1}$. Middle scoring since the defect does not reach 40% mass in the interval $[f_1; f_2]$.

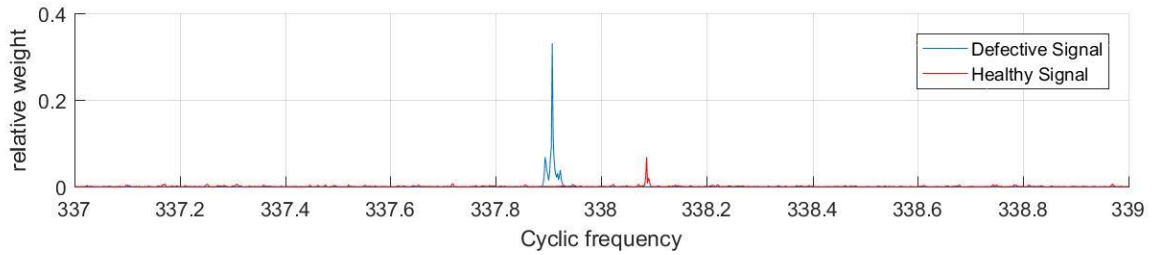


Figure 6: HPT 4 BPFO

It is clear that the arbitrary choice of $[f_1; f_2]$ presents an issue, especially since the characteristic frequency, which is not precisely known a priori, could be out of this small cyclic frequency span. To deepen this problem, figure 7 presents the gathered results obtained when several HPS are computed. The frequency span is the only varying parameter, and will be successively be shifted from 2 to 22 step by step (one step is $0.2 (lss)^{-1}$). The upper plot present the most probable characteristic frequency for each HPS, while the bottom plot present the probability of this “max frequency”. On the defective signal, the max frequency does not move from its first estimation, which happen to be the good one. On the contrary, the healthy signal reveals successively different frequencies ($338.1 (lss)^{-1}$ when $\Delta f \in [2; 3]$, $340 (lss)^{-1}$ when $\Delta f \in [3; 6.5]$, $341 (lss)^{-1}$ when $\Delta f \in [7; 16]$,...).

Moreover, in the case of the defective signal only, the probability associated to the most favoured frequency does not decrease noticeably along with the frequency span. This means that no other frequency channel appears within the extended interval able to compete with the BPFO cyclic frequency. On the contrary, the probability curve of the healthy signal is notably decreasing in magnitude, in a piecewise manner since every new local max introduces a small step increase of probability.

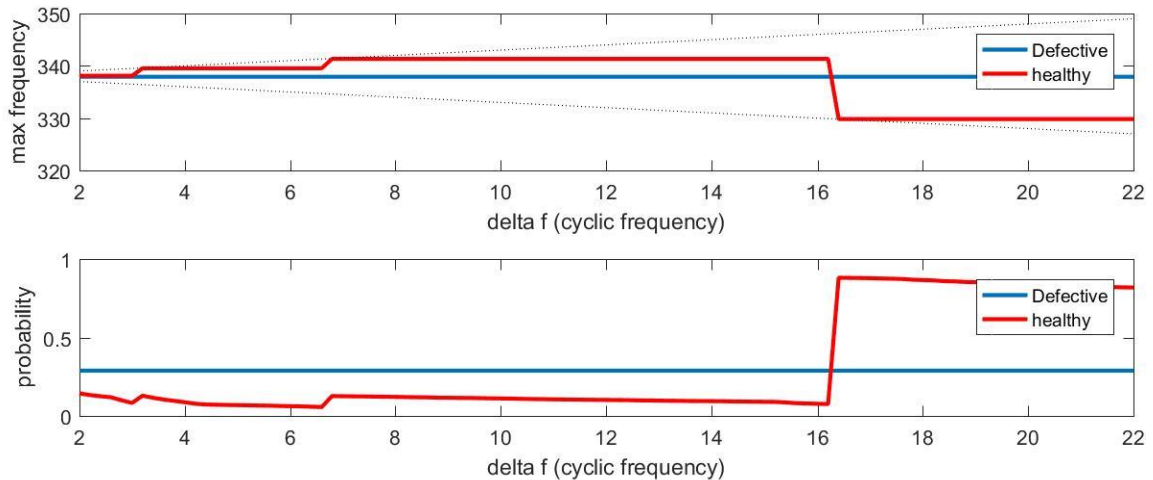


Figure 7: HPS local maxima depending on the frequency band of interest bandwidth. Top figure: frequency of the local maximum. Bottom figure: probability / relative weight of the local maximum. HPS computed with 8 harmo (K=8) and focused around BPFO.

Figure 8 present the same result but focused on BPFI frequency. The frequency identified by the HPS when $\Delta f \in [2; 4.5]$, is 506.84 and corresponds to the characteristic frequency observed from the high speed shaft encoder. However, this frequency is quickly overtaken by parasite phenomena, as soon as the frequency interval of interest is greater than $4.5 (lss)^{-1}$.

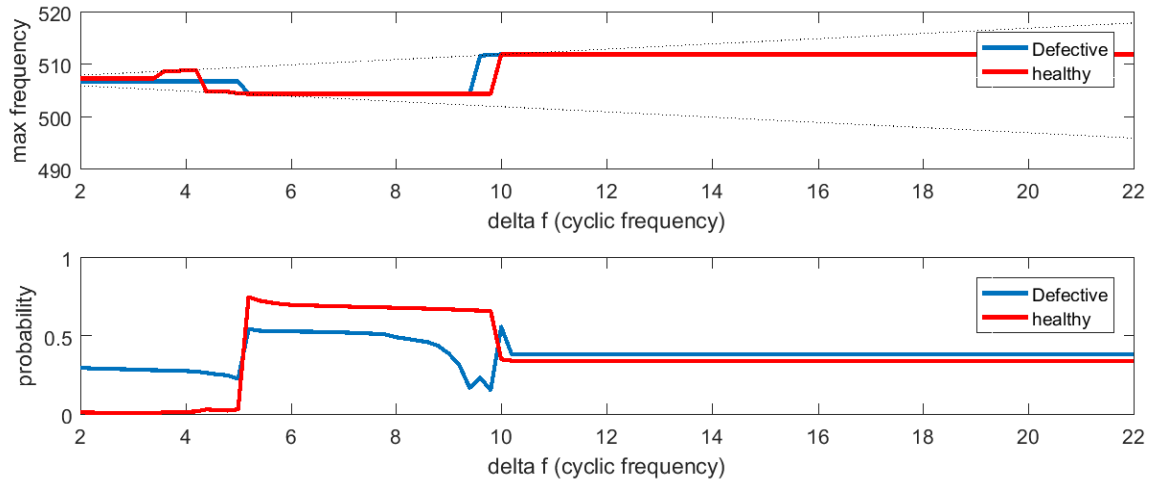


Figure 8: HPS local maxima depending on the frequency band of interest bandwidth. Top figure: frequency of the local maximum. Bottom figure: probability / relative weight of the local maximum. HPS computed with 8 harmo (K=8) and focused around BPFI.

4.3. MHPT

Figure 9 finally presents the results for the inner ring defect obtained with MHPS in a similar manner. The same frequency channel $f = 506.84(lss)^{-1}$ is strongly identified (~100% !) by this improve method and is not exceeded until the frequency band of interest is polluted by $f = 512(lss)^{-1}$. This annoying cyclic frequency corresponds to the interpolation frequency of the encoder, which is extremely energetic since this is a magnetic encoder. This problem has already been detailed in the literature, and could have been avoided by a proper filtering but has been kept to illustrate the limit of the proposed approach. If the parasite phenomena are not removed prior to HPS or MHPS processing, the expert is exposed to such a pollution. However, MHPS brings a clear improvement regarding modulated fault.

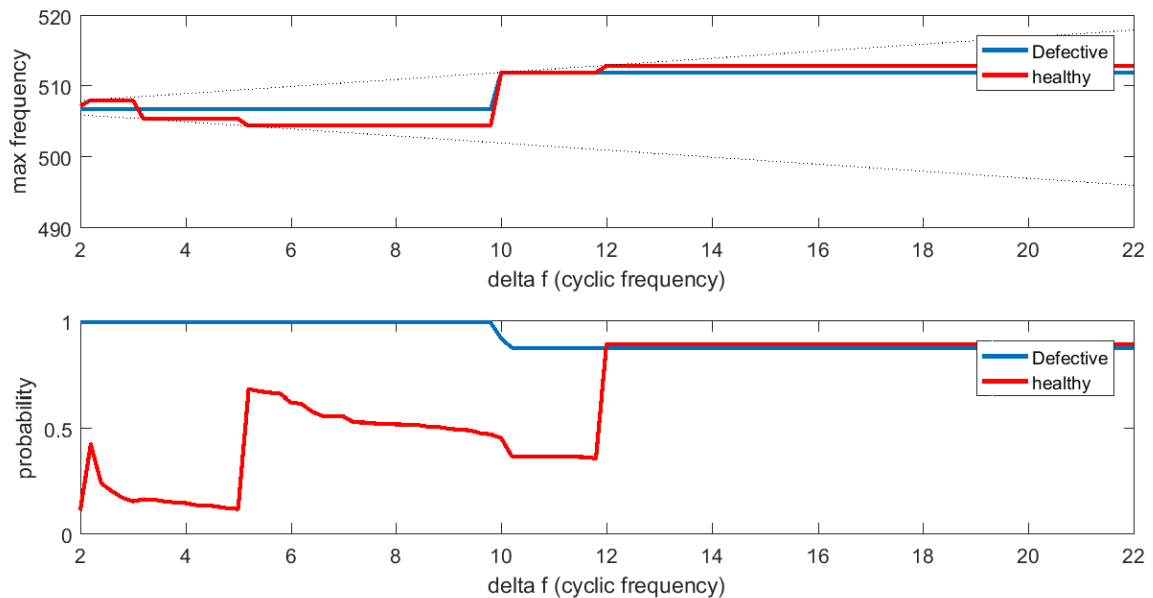


Figure 9: MHPS local maxima depending on the frequency band of interest bandwidth. Top figure: frequency of the maximum. Bottom figure: probability of the maximum. HPS computed with 4 harmos and 2 sidebands (K=4, M=2) and focused around BPFI.

5. CONCLUSION

This study shows the adaptation of HPS and MHPS to IAS signal applied to rotating machine monitoring. These “new tools”, only adapted from very old processing techniques discovered in the 60’s in the domain of speech analysis to precisely track voice pitch, help IAS monitoring to step forward in the Condition-Monitoring-System hall of fame ! The precise detection of an early bearing fault is now proven to be possible far away from the encoder location. In this paper, signal information was indeed shown able to step through the whole wind turbine shaft line. These tools present the ability to concatenate information from several elements of a set of harmonics, possibly modulated through a cyclic frequency, and therefore yields a more precise estimation of the characteristic frequency. This appears to be valuable in regards with two observations:

1. Vibration signals as well as IAS signals can contain a multitude of cyclic phenomena, especially since the whole shaft line can be observed from an encoder adroitly located. Once a bandwidth is determined by the expert as surrounding a characteristic frequency of interest, HPS or MHPS can respectively be used to clean this interval from either non harmonic content or non-modulated content.
2. Up to now, experts can hardly predict which harmonic number of a wide set makes the biggest contribution into the fault severity. Not only the diagnostic is simplified since the information is condensed in the characteristic frequency, but it is not needed anymore to follow simultaneously the various harmonics.

In the meantime, it is important to underline the drawbacks which were not overcome yet by the authors (damn them)... they are mainly linked to very strong hypothesis made at the beginning: the observed signal is expected to have one and only one set of harmonics.

1. If this is a real life signal, the chance is very few that only one set of harmonic set inhabits it. The expert must then be aware that the larger the frequency bandwidth, the smaller the harmonic set of interest, until it might not be prominent anymore. The proposal of the author to window the observation is suboptimal, since it increases the number of parameters to tune.
2. Going too far the other way, the risk that no set of harmonic lays within the frequency interval is also problematic. As was observed in this paper with healthy signal, it seems there is always a combination of frequency channel leading to something heavier than the average.

REFERENCES

- [1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, D. Rémond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, *Mechanical Systems and Signal Processing*, Volume 24, Issue 7, October 2010, Pages 1998-2011, ISSN 0888-3270, <http://dx.doi.org/10.1016/j.ymssp.2010.05.005>.
- [2] H. André, F. Girardin, A. Bourdon, J. Antoni, D. Rémond, Precision of the IAS monitoring system based on the elapsed time method in the spectral domain, *Mechanical Systems and Signal Processing*, Volume 44, Issues 1-2, 20 February 2014, Pages 14-30, ISSN 0888-3270, <http://dx.doi.org/10.1016/j.ymssp.2013.06.020>.
- [3] M.R. Schroeder, period histogram and product spectrum: new methods for fundamental frequency measurement, *J. Acoust. Soc. Am.* 43 (1968) 829-834.
- [4] A.M. Noll, Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate, In *Proceedings of the Symposium on Computer Processing Communications*, 1969 Pages. 779-797.
- [5] H. André, A. Bourdon, D. Rémond, On the use of the Instantaneous Angular Speed measurement in non-stationary mechanism monitoring. *Proceedings of the ASME 2011 International Design Engineering Technical Conferences & 23rd Biennial Conference on Mechanical Vibration and Noise, IDETC/CIE 2011*, Washington, DC, USA
- [6] P. D. McFadden, J. D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound Vibration*, vol. 96, 1984, p. 69-82.
- [7] J.L. Gomez, A. Bourdon, H. André, Didier Rémond, Modelling deep groove ball bearing localized defects inducing instantaneous angular speed variations, *Tribology International*, Volume 98, 2016, pp 270-281

A Comparison of Architectures and Evaluation of Metrics for In-Stream Machine Learning Algorithms in Industry 4.0 Applications.

C. Berry, G. Hall, B. Matuszewski, L.K. Shark

Advanced Digital Manufacturing Technology Research Centre, University of Central Lancashire, Preston
PR1 2HE, UK

ABSTRACT

Modern manufacturing processes can be monitored using a wide variety of sensors that can be used for tasks such as process control, predictive maintenance, and defect detection and prediction. It would be desirable within an Industry 4.0 framework to use machine learning techniques to both optimise data collection and develop pattern recognition. In order to accomplish this the application must be able to handle (near) real-time processing using streaming data obtained directly from the manufacturing process along with other appropriate sources. In order to update and evaluate the main streaming channel of the framework we propose a second longer term channel that can evaluate and optimise competing algorithms and data strategies to select the most appropriate for in-line detection or control. This paper looks at Lambda and Kappa architectures for Big Data analysis and methods of evaluating between competing in-stream machine learning techniques and data selection criteria as part of a dual channel Industry 4.0 intelligent system. The paper will examine criteria and metrics for selection and evaluation of suitable streaming learning algorithms along with architecture and strategy selections for live switching and modification of in-process control and monitoring processes.

Keywords: Big Data, Streaming Analytics, Industry 4.0, Lambda Architecture, Machine Learning

Corresponding author: C. Berry (email: cberry1@uclan.ac.uk)

1. INTRODUCTION

Recent advances in engineering and computing areas have led to the creation of Industry 4.0, a combination of related fields such as cloud computing, large scale analytics (Big Data), the Internet of Things (IoT) along with automation and robotics. Industry 4.0 provides many opportunities for manufacturing companies to process and look at the data around their manufacturing processes in far closer detail than ever before but this also presents many challenges. In this paper we will look at some of the challenges related to Industry 4.0 applications and will present a proposal for a system that intelligently monitors and/or controls a manufacturing process whilst simultaneously evaluating and, if necessary replacing, the algorithms it uses to carry out the monitoring and control process.

In order to accomplish this type of operation there are a number of potential issues that we need to consider:

- The nature, speed and amount of data being collected by the manufacturing system.
- The architecture and framework used to process the data.
- Algorithm selection for control and monitoring.
- Selection criteria and metrics for evaluating competing machine learning algorithms.

2. DATA AND ARCHITECTURE.

2.1 Data Acquisition.

In an Industry 4.0 application the opportunity is available to collect large amounts of in-process data for use with Big Data style large scale analytical processing. In a system where the data is being stored off-line for later processing this can be achieved with relative ease [6], data can be collected via a large number of sensors and transmitted back to a central point to be saved and later analysed. In this scenario provided appropriate data collection is carried out, suitable network capacity and cluster/server capacity and throughput are available along with sufficient storage, the problem is an up-scaled version of traditional data processing carried out by many intelligent manufacturing processes and has been visited a number of times before[7],[8]. In the case where the application is being used for control the data must be collected, analysed and acted upon in real-time (or near real-time). This presents further complications that must be considered.

In a streaming application data must be processed quickly enough for the results not to become obsolete before the calculations have been completed. In control processes lagging data can have seriously consequences resulting in system degradation or complete failure. In the Big Data ecosystem three processing engines have emerged as the main focus for processing streaming data, these are; Apache Spark Streaming [9], Apache Storm [10] and Apache Flink [11]. Each of these processing engines has its own set of features and a full examination of these would be beyond the scope of this paper, we will limit the discussion here to the main features that concern our implementation and direct the reader to more complete surveys in [12] and [13]. In order to select the preferred processing engine however we must first consider the software framework under which it would be running.

2.2 Software Framework.

The proposed software architecture is shown in figure 1 and will be discussed below. There are a number of competing protocols within the IoT arena but for many applications MQTT is emerging as the de facto standard [14] and has attained an ISO/IEC release as a standard [15], as such this will be the protocol that we will assume sensor data is being transmitted in. In all instances the processing framework would be running under Hadoop running YARN [16] for our implementation, this is by far the most common Big Data platform. The first problem to overcome is getting the MQTT data into Hadoop HDFS (Hadoop Distributed File System), for this we can use Apache Kafka [17] via a MQTT to Kafka bridge, of which there are a number available such as [18] and [19]. Kafka is a message / subscriber system that creates real-time data pipelines between applications and can be used to send data to different destinations, in the case of our implementation it is used to store data for long term analysis in a central storage database and also to send the data to our real-time processing engine.

It is the goal of our system to be able to intelligently select between competing machine learning algorithms and implement the preferred algorithm in our real-time control channel. In order to do this we need to consider frameworks that allow the system to perform long term analysis of the data whilst simultaneously running a real-time processing channel. Such a system was described by Nathan Marz in [1] when he coined the term Lambda architecture as shown in figure 2 (from [1]), this architecture has also been covered in [2]. In this type of architecture there are two processing channels, one takes streaming data directly from the source and processes it in real-time, the second channel stores data in permanent storage and uses this for batch processing to find long term patterns or trends that can later be used to adjust the first real-time channel. It is our goal to have this long-term channel make autonomous decisions regarding adjusting and ultimately replacing the real-time algorithm without the need for a human data analyst. In the standard Lambda model although the two channels run simultaneously there is no direct link between the two channels and it relies on human data analysts to adjust the real-time stream based on the results of the long-term batch channel. The Lambda architecture is not without its critics, Kreps in [3] argued that the Lambda architecture was problematic in nature due to the need to write and maintain two separate code bases. Typically the real-time channel of a Lambda architecture would be written using a

streaming processing engine such as Storm or Spark Streaming and the long term batch analysis would be done using MapReduce, this is due to code being written to match the architecture it is running on, to get the best out of a streaming algorithm it has to be written for streaming not for batch. With distributed data processing applications being difficult to write and maintain in the first place Kreps argues that it is unreasonable to expect a company or organisation to support and debug to entirely different code bases. In place of the Lambda architecture Kreps proposes a Kappa architecture, see figure 3 (from [3])

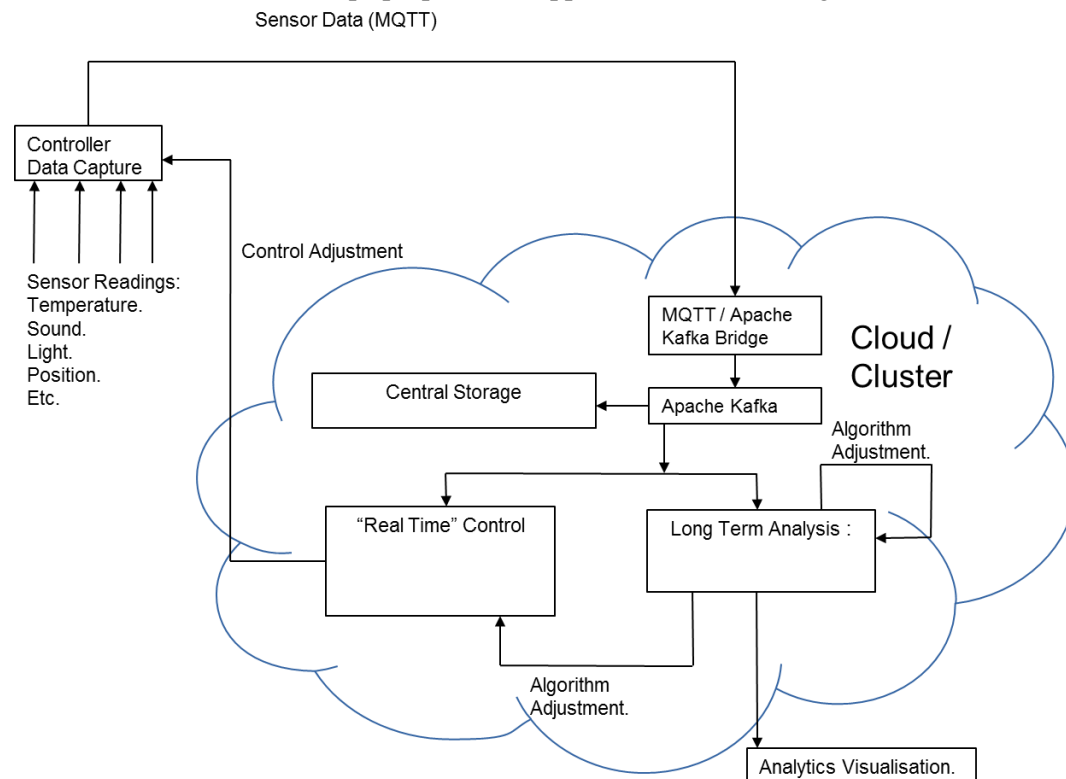


Figure 1 – Proposed Software Architecture

The Kappa architecture uses Kafka's log system to store data for an amount of time (Kafka can store very large amounts of data in log files, Kreps gives the example that LinkedIn stores more than a petabyte of data in Kafka logs on-line), if another process wishes to perform analysis alongside the real-time stream then another stream is opened taking from the beginning (or elsewhere) of the Kafka log. Separate streams may write data into HDFS storage but this is not considered to be used as part of the processing system. The Kappa architecture differs from the Lambda architecture in that for Lambda there is always a dedicated long term channel whereas Kappa spawns a new streaming channel as needed for reprocessing tasks.

Part of the problem directed at the Lambda architecture is that of multiple code bases, particularly as applications written for MapReduce and streaming engines are very different, this is what we have to bare in mind when choosing our processing engines. Returning to the selection of our processing engine our selection here is guided by the previous discussion on Lambda and Kappa along with the specifics of a manufacturing system.

Apache Storm is a true in-memory streaming process engine that can process single event data as it arrives, this would make it a good choice for the real-time channel of our architecture, however this would mean using a different engine in the long-term analysis channel if we wanted to process data in batch mode. We could adopt the Kappa model of running multiple Storm channels for reprocessing previous data. An alternative is to use Apache Spark Streaming, Spark is an in-memory processing engine and Spark Streaming extends the Spark engine to handle near real-time streams. A choice of Spark Streaming would allow us to run Spark Streaming in the real-time channel and Spark in the long-term channel running in batch mode. Spark Streaming, however, is not a true streaming engine, it processes data in micro batches,

that is small bundles of data, but it does not do true “as it arrives” stream processing. In many cases Spark Streaming’s micro batch processes quickly enough as to not be an issue but in the case of a real-time control system on a manufacturing process the latency issues could cause severe problems.

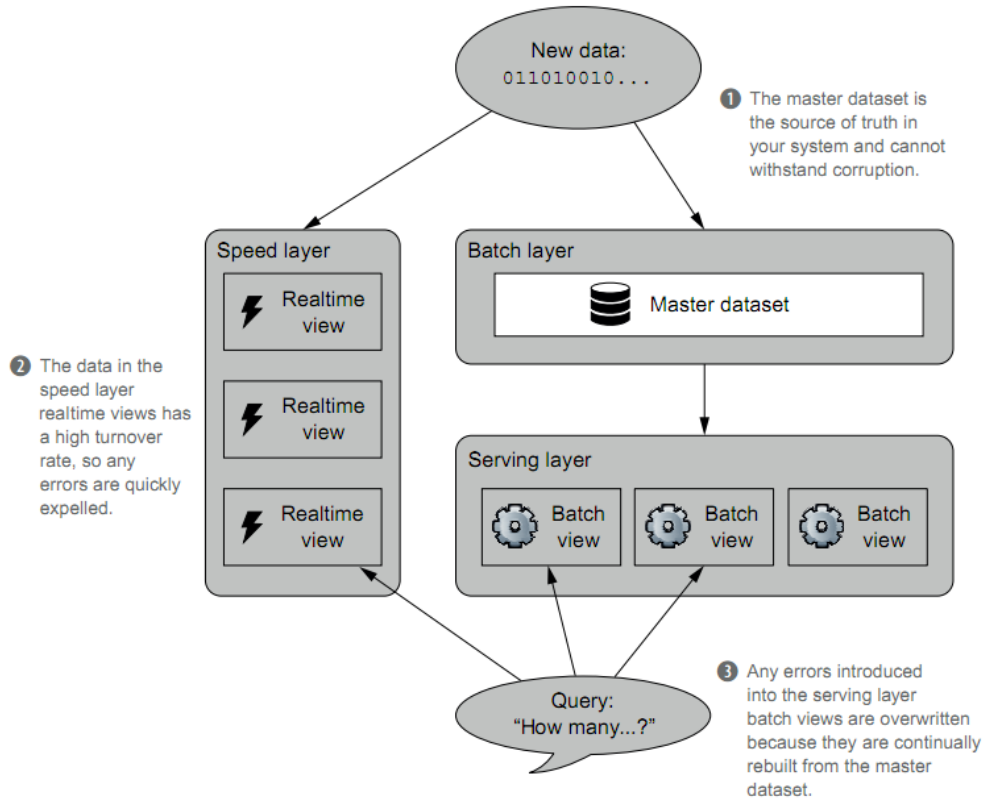


Figure 2 - Lambda Architecture (from [1])

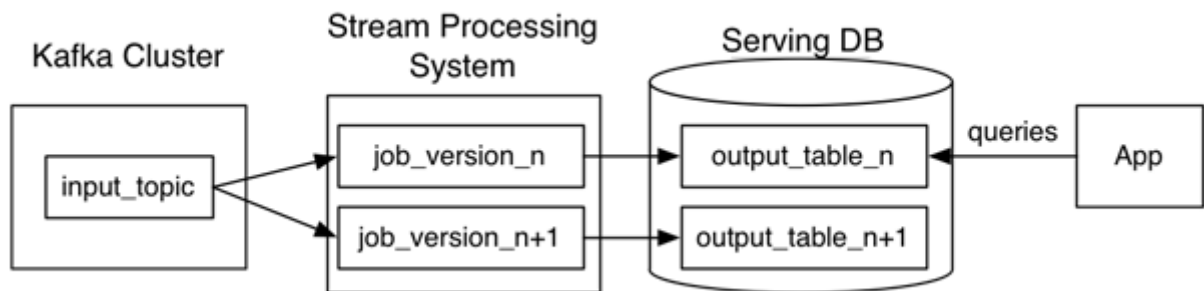


Figure 3 – Kappa Architecture (from [3])

Apache Flink is a relatively new in-memory processing engine that can process data in either batch or streaming modes, this would allow us to use the same processing engine for both the real-time and long-term batch channels. The disadvantage here with Flink is that as a relatively new engine the library support for add-ons such as machine learning libraries is not as well established as it is for the more mature Spark or Storm.

We have made the decision in our case to go with an Apache Flink implementation due to its dual nature of stream and batch processing. Although Storm via Kappa is a valid option here we felt that having the flexibility of introducing new datasets at the testing stage by accessing off-line databases through HDFS was useful enough to balance the extra effort involved in running a Lambda Flink architecture with a reduced set of available libraries. In the case here, our long-term term processing channel will have to

simulate, at least in part, a streaming application, so that an evaluation can be made regarding its potential use in the real-time channel. It is therefore advantageous to have both channels using the same processing engine to minimise differences in performance due to changes in architecture.

2.3 Machine Learning Algorithms.

Modern manufacturing has a wide variety of processes that could benefit from large scale data analytics in this implementation we have chosen to look at the application of predictive analysis/modelling. To this end we have selected two algorithms in the initial testing phase based on the popularity of these algorithms in the field [20] these are random forests and elastic net. These two algorithms have been used successfully in manufacturing and industrial applications to predict a wide range of functions such as tool wear [21], electrical discharge machining equipment [22], and semi-conductor production [23]. The Flink machine learning library (FlinkML) is a collection of machine learning algorithms that have been prewritten for use within a Flink framework, both elastic net and random forests are on the roadmap for implementation of this library [24], meaning that they have not yet been added to the library, this means that this project will have to implement these algorithms and this is currently the main focus on ongoing work. The relative immaturity of Flink means that the library support is not as robust yet as for an engine such as Spark which has its own machine learning library with support for random forests and Lasso and Ridge regression (and from there elastic net would be an easier implementation), however as discussed in section 3., we feel the trade-off will be beneficial in the long run.

3. EVALUATION METRICS.

At any given time in the manufacturing process, depending on the data being collected or the process being observed, there is no guarantee that the particular machine learning algorithm selected at a previous time is the most optimal one to be currently running. As metrics change and the observed components of the system are altered or wear we may find that another approach becomes more suitable. The question we need to address is what do we mean by terms like suitable and optimal? There is a danger when comparing machine learning algorithms to focus on accuracy as the final word on which performs best but in a number of cases this is not the only or the most important factor.

In order to properly compare algorithm implementations there are two aspects that need to be considered, the accuracy of the algorithms and the efficiency of the algorithms with regard to being able to make timely decisions. [4] looked at a wide variety of machine learning algorithms and established a number of criteria for determining their accuracy. In this study the authors identified 8 accuracy measures ; three threshold metrics (accuracy, F-score and lift), three ordering / ranking metrics (area under ROC curve, average precision, precision/recall break even point) and two probability metrics (squared error and cross-entropy). Multiple measures are required to assess an algorithms relative accuracy as depending on what the algorithm is used for we are likely to be interested in measures beyond simple accuracy of how many of the training example an algorithm gets right. The other measures in the tests track metrics such as bias towards a particular outcome, relative valuing of true positives, false negatives, true negatives and false positives which will be application specific to the process being monitored.

Although the accuracy element is an important part of the algorithm evaluation, in this project we also have the problem that the algorithm will have to cope with the streaming nature of the data. As described earlier, in a Lambda architecture there is a streaming element along with a static batch element. In this project the streaming element is selected via the batch element, so although the batch element is static and has access to the full dataset at the beginning of training it must behave as a streaming algorithm in order to fully simulate the environment under which the tested algorithm will run, therefore the batch testing here will be unable to take full advantage of being a static process. The additional challenges of evaluating a streaming algorithm are explored in [25] and [5]. According to [5] influences on streaming data learning algorithms are; memory space, due to streaming algorithms having to take place in-memory due to speed requirements, learning time, the algorithm needs to be able to process and learn from incoming samples as they arrive and generalisation power, generally how good is the model at capturing the underlying trends and patterns in the

data. In selecting an optimised algorithm in the long-term trends channel this project will take into consideration the 8 accuracy metrics from [4] along with a measure of processing speed compared to the speed of incoming data, the algorithm will need to be able to produce results in a timely manner compared to the incoming frequency of the data stream. Finally another metric that needs to be taken into consideration is the life expectancy of the new algorithm. We need to be confident that the new algorithm is not going to outperform the old one for a short amount of time before needing to be swapped back otherwise we run the risk of spending most of the architectures time and resources on swapping between algorithms, for this we suggest introducing a minimum time frame that the algorithm needs to be predictably superior for. This minimum time frame would need to be set with regard to the length of time needed to make the change over and the effect it has on the system overall, particularly if it requires different sensor readings.

By choosing Apache Flink as the processing engine for the two channels we hope to avoid to as large extent as possible the problems raised by Kreps [3] regarding the multiple code bases. Also by using the same engine in both the real-time and long-term channels we hope to ease one of the other problems faced by the project, that of swapping out the machine learning algorithm. Once the long-term engine has found a better algorithm or a better set of input data for the current algorithm in the real-time channel we then have the problem of swapping out the algorithm in real-time. For a system running two different engines this would present a real problem, if a MapReduce application decided that a random tree algorithm was now more appropriate than a elastic net algorithm could the system be certain that this would remain the case once the algorithms were running in real-time mode using Storm or Spark Streaming? This problem is lessened by running both channels with Apache Flink, by being able to test the stream characteristics in the long-term channel we can be hopeful that the algorithm will perform with the same efficiency and accuracy in the real-time channel. The process of swapping over the algorithms is another issue. In the case of a manufacturing system we cannot afford to simply stop the control system whilst changing out the algorithm nor would stopping production be an acceptable solution in most real manufacturing setups. This is where Big Data again shows its strength, because the Big Data ethos is for systems to be scalable and because Apache Kafka can supply data to any number of clients we can create a second control / real-time stream running in parallel to the original one. Once the new stream is up and running we can again perform in-line checks to make sure it is running with the improvements the long-term channel predicted and if this is the case then we can stop the original channel and allow the new one to take over. If we are running with limited resources through either economical restrictions on cloud use or because the architecture is running on an in-house cluster then we could for the lifetime of the change-over shut down the long-term channel and allocate its resources to the new real-time channel, using the old real-time's recovered resources to restart the long-term channel once change-over has occurred.

4. CONCLUSIONS AND FUTURE WORK.

The architecture presented in this paper is a blueprint for an autonomous Lambda architecture based on manufacturing processes taking predictive modelling as a first application. Work has started on adapting two chosen algorithms, that have previously been successful in streaming frameworks, to work with this architecture, at the same time we have looked at a number of evaluation metrics for being able to make an informed and sensible choice between two competing machine learning algorithms. The project aims to potentially remove the role of data analyst from such systems by allowing the system itself to make decisions regarding the best algorithm to control that process at any given time. There are a number of challenges ahead particularly the process of swapping algorithms in real-time, this paper has discussed the idea of scaling the hardware base after the selection process and during the change-over period to allow for parallel running of the two algorithms to allow for a smooth, seamless swap between the two algorithms. In terms of future work, once the initial architecture is running and we have reported findings on the initial selected algorithms, we hope to expand the number of algorithms being considered, at this time the leading contender for the next stage would be to try a Deep Belief Network (DBN) as these have recently shown strong results in predictive systems [26], [27].

REFERENCES

- [1] Nathan Marz with James Warren, Big Data – Principles and Best Practices of Scalable Real-Time Data Systems, Manning Publications, 2015.
- [2] Mariam Kiran, Peter Murphy, Inder Monga, Jon Dugan and Sartaj Singh Baveja, Lambda Architecture for Cost-effective Batch and Speed Big Data Processing, 2015 IEEE International Conference on Big Data.
- [3] Jay Kreps, Questioning the Lambda Architecture, O'Reilly Radar Blog, July 2, 2014.
<https://www.oreilly.com/ideas/questioning-the-lambda-architecture>
- [4] Rich Caruana and Alexandru Niculescu-Mizil, An Empirical Comparison of Supervised Learning Algorithms, Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, 2006.
- [5] Joao Gama, Raquel Sebastiao and Pedro Pereira Rodrigues, On evaluating stream learning algorithms, Machine Learning (2013) 90:317–346.
- [6] Xuhua Zhang, Qiaoyan Wen., A Policy Anomaly Detecting Algorithm Based on MapReduce, International Conference on System Science, Engineering Design and Manufacturing Informatization, 2011.
- [7] Salim Charaniya, Huong Le, Huzefa Rangwala, Keri Mills, Kevin Johnson, George Karypis, Wei-Shou Hu, Mining manufacturing data for discovery of high productivity process characteristics, Journal of Biotechnology, 147 (2010) 186–197.
- [8] A. K. Choudhary, J. A. Harding, M. K. Tiwari, Data mining in manufacturing: a review based on the kind of knowledge, J Intell Manuf (2009) 20:501–521.
- [9] Apache Foundation, Apache Spark Streaming homepage, <https://spark.apache.org/streaming/>, accessed 19/03/2017.
- [10] Apache Foundation, Apache Storm homepage, <https://storm.apache.org>, accessed 19/03/2017.
- [11] Apache Foundation, Apache Flink homepage, <https://flink.apache.org>, accessed 19/03/2017.
- [12] Sanket Chintapalli, Derek Dagit, Bobby Evans, Reza Farivar, Thomas Graves, Mark Holderbaugh, Zhou Liu, Kyle Nusbaum, Kishorkumar Patil, Boyang Jerry Peng and Paul Poulosky. Benchmarking Streaming Computation Engines : Storm, Flink and Spark Streaming, 2016 IEEE International Parallel and Distributed Processing Symposium Workshops.
- [13] Guenter Hesse and Martin Lorenz, Conceptual Survey on Data Stream Processing Systems, 2015 IEEE 21st International Conference on Parallel and Distributed Systems.
- [14] Sharon Shea, TechTarget.com, IoT Messaging : The MQTT protocol is stepping up to the plate,
<http://internetofthingsagenda.techtarget.com/feature/IoT-messaging-The-MQTT-protocol-is-stepping-up-to-the-plate>, last updated 23/02/2016, accessed 19/03/17.
- [15] prweb.com, OASIS MQTT Internet of Things Standard Now Approved by ISO/IEC JTC1,
<http://www.prweb.com/releases/2016/07/prweb13562319.htm>, last updated 19/07/2016, accessed 19/03/2017
- [16] Apache Foundation, <https://hadoop.apache.org/docs/r2.7.3/hadoop-yarn/hadoop-yarn-site/YARN.html>, accessed 19/03/2017.
- [17] – Apache Foundation, <https://kafka.apache.org/>, accessed 23/03/2017.
- [18] Dhruv Kalaria, Bridge MQTT Messages from Mosquitto to Kafka Broker on the same topic,
<https://github.com/DhruvKalaria/MQTTKafkaBridge>, last updated 26/11/2016, accessed 19/03/2017.
- [19] iotresearcher, MQTT Kafka bridge, <https://github.com/iotresearcher/mqtt-kafka-bridge>, last updated 21/10/2016, accessed 19/03/2017.
- [20] Jeremy Howard, Mike Bowles, The two most important algorithms in predictive modeling today, O'Reilly Starta conference 2012, <http://conferences.oreilly.com/strata/strata2012/public/schedule/detail/22658>, accessed 20/03/2017.
- [21] Dazhing Wu, Connor Jennings, Janis Terpeny, Soundar Kumara, Cloud Based Machine Learning for Predictive Analytics: Tool Wear Prediction in Milling.
- [22] Indramil Ghosh, M.K. Sangal, R.K. Janu, Pranab K. Dan, Machine Learning for Predictive Modeling in Management of Operations of EDM Equipment Product, 2nd International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2016.
- [23] Gian Anronia Susto, Adrian B. Johnston, Paul G. O'Hara and Sean McLoone, Virtual Metrology Enabled Early Stage Prediction for Enhanced Control of Multi-Stage Fabrication Process, IEEE International Conference on Automation Science and Engineering (CASE), 2013.
- [24] Theodore Vasiloudis, FlinkML: Vision and Roadmap,
<https://cwiki.apache.org/confluence/display/FLINK/FlinkML%3A+Vision+and+Roadmap>, last updated 27/7/2016, accessed 21/03/2017.
- [25] Nigel Williams, Sebastian Zander, Grenville Armitage, Centre for Advanced Internet Architectures (CAIA) Technical Report 060410B, Swinburne University of Technology.
- [26] Yubo Tao and HongKun Chen, A hybrid wind power prediction method, Power and Energy Society General Meeting (PESGM) 2016.
- [27] Yuhua Jia, Jianping Wu and Yiman Du, Traffic speed prediction using deep learning method, 9th IEEE conference on Intelligent Transportation Systems (ITSC) 2016.

Measurement Uncertainty When Measuring Thermal Effects on a Machine Tool Using Non-Contact Displacement Transducers

S Gorton¹, S Fletcher¹, D Clough², A P Longstaff¹

¹Centre for Precision Technologies, School of Computing & Engineering, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK

²Machine Tool Technologies Ltd, 1H Ribble Court, 1 Meadway, Shuttleworth Mead Business Park, Padiham, BB12 7NG, UK

ABSTRACT

Manufacturing to tight tolerances on a machine tool puts a greater emphasis on understanding and quantifying the errors associated within the system. The largest contributors to errors in many machine tools are likely to be related to thermal effects. Non-contact displacement transducers (NCDT) are a suggested measurement sensor under the ISO230 series of standards which include the test code for machine tools. A measurement uncertainty budget that identifies all known sources of errors should be performed before conducting a test. However, the complexities of a machine tool and long test duration require that some uncertainties are quantified using a type B evaluation which can lead to under or overestimating the different errors.

This paper will look at the uncertainty budget for using an NCDT system for measuring thermal displacement in a machine tool, and the benefits of the system being employed as part of a service package in accordance with ISO230-3.

Keywords: Machine tool, End user applications, Standardisation, Thermal effects

Corresponding author: Mr Sean Gorton (sean.gorton@hud.ac.uk)

1. INTRODUCTION

Conducting a robust and traceable measurement uncertainty budget is critical to being able to understand and improve the accuracy of machining processes. The errors caused by thermal deformation directly affect the precision of a machine tool [4], and due to the complexity of a machine tool they are notoriously difficult and time-consuming to assess. By employing, a robust set of procedures and processes, the repeatability and accuracy of the measurement can be maintained to provide well-informed data to the customer.

1.1. Standards

There are three main tests for assessing the thermal effects on a machine tool under BS ISO 230-3:2007 they are:

- Thermal Distortion Caused by Rotating Spindle (spindle heating)
- Environmental Temperature Variation Error (ETVE) test
- Thermal Distortion Caused by Linear Motion of Components (axis heating)

The spindle heating test can be performed to assess the thermal effects associated with rotating the spindle and how the heat flows through the structure causing thermal distortion between the machine table (workpiece) and the tool. An ETVE test can be performed to assess how the environmental temperature fluctuations effects the accuracy. An axis heating test can be carried out to identify the effects of the heat

generated from components such as the balls screws, nut and support bearings and linear axis drive motors impact on the thermal distortion of the machine tool [1].

Non-contact electronic sensors are one of the few linear displacement sensors that are suggested to be used to conduct the thermal tests. The ISO 230-2 standard provides guidance for performing a measurement uncertainty estimation for linear position measurement but no evaluation has been performed for these thermal tests.

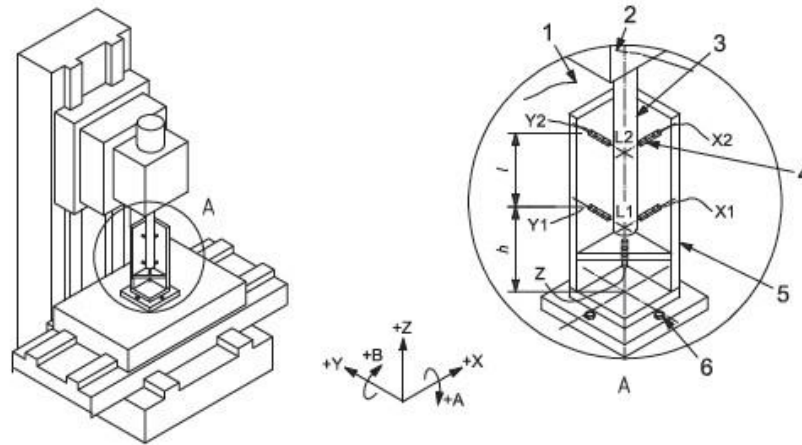


Figure 1. Typical set up for tests conducting ETVE test, thermal distortion caused by rotating spindle and by moving linear axis [1]

2. MEASUREMENT UNCERTAINTY EVALUATION OF SUPPLIED SENSOR(S)

When outsourcing a sensor with a calibration certificate, the traceability of how the accuracy of the sensor was determined can be difficult if not impossible to obtain from the supplier. Therefore, by having an understanding of the variables that will contribute to the accuracy of the system is important in ensuring the sensor and system are fit for purpose.

2.1. Review of sensor capability, specification and calibration certificate

The NCDT chosen for this work is an eddyNCDT DT3005 sensor which provides the requisite accuracy and dynamic range for thermal tests. In contrast to the commercial system from Lion Precision, which uses capacitance based NCDTs, these inductive sensors have the added benefit of robustness against contamination in the interfacial fluid (air) between the sensor and the target object being measured. The lower accuracy due, for example, to the electrical run-out is averaged out for thermal tests [2].

The linearity of the NCDT is an important parameter to be understood over a long period of time because it is directly related to the accuracy of the sensor. The temperature stability and the temperature compensation capability of the sensor and controller also needs to be understood over a large period to ensure a good understanding of how the sensor performs whilst providing traceability. The specification highlights the temperature stability of the sensor at a voltage output corresponding to the midrange but how does the sensor perform at the furthest distance from the target? This is only one of numerous questions that need to be understood.

Table 1: Typical specification of NCDT

| Variables | Non-contact displacement transducer |
|-------------------------------------------|-----------------------------------------------------------------------------|
| Measurement object material | Steel (ST37 DIN10037) |
| Measurement range | 1mm |
| Offset distance | 0.1mm |
| Linearity | $\leq 0.25\%$ full scale output (FSO) $2.5\mu\text{m}$ |
| Resolution | $\leq 0.05\%$ full scale output (FSO) $0.5\mu\text{m}$ |
| Repeatability | $\leq 0.05\%$ full scale output (FSO) $0.5\mu\text{m}$ |
| Max sensitivity range | $\leq 1\%$ full scale output (FSO) $10.0\mu\text{m}$ |
| Bandwidth | 5kHz(-3dB) |
| Temperature stability (midrange - MMR) | 0.025% FSO / $^{\circ}\text{C}$ – $0.25\mu\text{m}$ / $^{\circ}\text{C}$ |
| Temperature compensation range sensor | 10°C - 125°C |
| Temperature compensation range controller | 10°C - 60°C |
| Ambient temperature range | -30°C - 125°C |
| Design | unshielded |
| Recommended object geometry (flat) | $\varnothing 24$ (4 times diameter of sensor) |
| Sensor cable length | 1m |
| Output | 0.5V – 9.5V |
| Power supply | 12V – 32V |
| Protection class | IP67 |
| Pressure resistance | 10 bar (sensor cable and controller) |
| Weight | 70g |

Figure 2 shows a straight line graph created in excel from data taken from the supplier's calibration certificate. The calibration certificate states that the mentioned measuring system was calibrated in compliance with an accredited quality management system which has been certified to DIN EN ISO 9001:2008. The used measuring devices are regularly calibrated and traceable through reference standards to recognized national standard laboratories in accordance to the International Unity System. The test report is valid for the reported system configuration at test environment of 25.2°C and under standard mounting of sensor. Unfortunately, the accuracy of the temperature sensor isn't stated in the calibration certificate and the definition of the 'standard mounting of the sensor' isn't included.

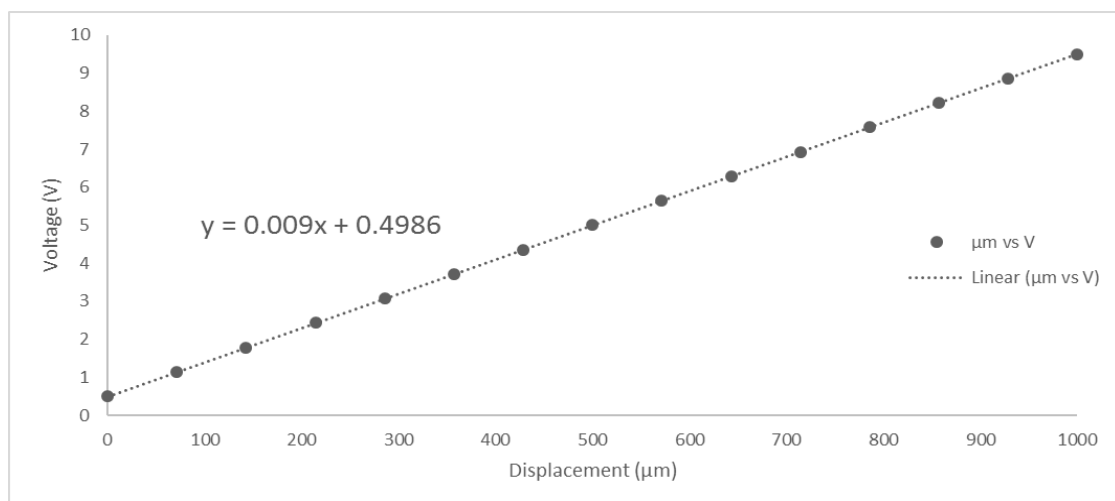


Figure 2 Straight-line graph for displacement vs voltage from supplied calibration certificate

The voltage gradient shown by the linear equation provided in figure 2 provides agreement with the brochure over the measurement range 0.5V to 9.5V showing a 1 μ m change in displacement corresponding to 0.009V.

The sensor specification states a linearity output of <2.5 μ m (see table 1). While this can be used in the uncertainty estimation, additional evaluations were performed to confirm this parameter and to quantify how temperature change had an influence on the other aspects of the system such as the sensor mounting and the signal conditioning electronics. Table 2 includes results from the evaluation of temperature effect on the sensor and electronics with the linear, cubic and quartic errors associated with creating a polynomial equation for the voltage fluctuation over the temperature range and a quadratic equation for the likely temperature range of the displacement sensor between 10°C and 30°C respectively. Also included in the table, the voltage output at the reference temperature of 20°C for the equations created so that the error could be evaluated from the reference temperature.

You can see from table 2 and table 3 Voltage fluctuation with temperature change of the electronics the linear error associated with the temperature of the sensor and electronics can result in an uncertainty of 3.256 μ m, which is obtained by adding the maximum linear errors of 1.136 μ m and 2.120 μ m. By understanding the application, it is unlikely that the test would be conducted outside a temperature range between 10°C and 30°C therefore by fitting a polynomial equation the uncertainty can be reduced to 0.045 μ m for thermal effects associated with the temperature sensor and to 0.016 μ m for the thermal effects on the electronics. This results in an uncertainty of 0.061 μ m.

Table 2: Voltage fluctuation with temperature change of the sensor and electronics

| Temperature sensor data (TSD) | | | | |
|----------------------------------------|-----------|---------------------------------------------------------------|---------------------------------------------|-------------------------------------------|
| °C | V | Linear error from linear equation in μm | Quadratic error in μm | Quartic error in μm |
| 20 calculated voltage (linear) | 4.946 | | | |
| 20 calculated voltage over application | 4.938 | | | |
| 20 calculated voltage (quartic) | 4.935 | | | |
| 125.3 | 4.946 | 0.024 | | 0.050 |
| 89.9 | 4.949 | -0.309 | | 0.046 |
| 39.8 | 4.947 | -0.087 | 0.045 | 0.045 |
| 21.7 | 4.936 | 1.136 | 0.044 | 0.044 |
| 9.8 | 4.954 | -0.864 | 0.045 | 0.045 |
| | max error | 1.136 | 0.045 | 0.050 |

Table 3: Voltage fluctuation with temperature change of the electronics

| Temperature electronics data (TED) | | | |
|-------------------------------------------|-----------|-----------------------------------------|------------------------------------------|
| °C | V | Cubic error in μm | Linear error in μm |
| 20.0 calculated voltage (cubic) | 4.954 | | |
| 20.0 calculated voltage (linear) | 4.963 | | |
| 60.2 | 4.960 | -0.053 | -2.120 |
| 34.9 | 4.952 | 0.010 | -1.884 |
| 21.8 | 4.954 | 0.016 | -1.080 |
| 9.7 | 4.959 | 0.011 | 0.013 |
| | max error | 0.016 | -2.120 |

2.2. Practical use of sensor

The practical application of the sensor needs to be understood to ensure that various variables are quantified and if possible removed from the testing to improve the accuracy of the measurement. The warm up cycle was conducted without any external influences such as the fixture design and test bar homogeneity. The warm up cycle was conducted over a one-hour period. Less than 1 μ m change from after 1 hour to 3 hours from the initial test was noted. The measurement target test was conducted to identify how the change in target impacts on the linearity of the voltage output.

Table 4: Warm up effect on uncertainty

| After warm up | Difference in output | Error in μ m |
|---------------|----------------------|------------------|
| 9.083 | 0.019 | 2.111 |
| 9.089 | 0.019 | 2.111 |

The measurement target (i.e. test bar) test was conducted to identify how the change in target affects the linearity of the voltage output. This is important when taking a measurement and providing accurate traceability between the target during calibration and the accuracy of the target during thermal testing.

Table 5: geometry target effect on sensor measurement uncertainty

| Target Type | Min | Max | Difference in V | Error in μ m |
|------------------------------|---------|---------|-----------------|------------------|
| Flat plate 1mm thick | -0.814 | -0.82 | -0.006 | -0.6666667 |
| 4mm diameter rod | -0.478 | -0.493 | -0.015 | -1.6666667 |
| 31mm diameter 1mm thick tube | -0.8736 | -0.8776 | -0.004 | -0.4444444 |
| 40mm diameter 1mm thick tube | -0.834 | -0.837 | -0.003 | -0.3333333 |

The warm up cycle can be removed from the uncertainty budget with the necessary knowledge and the sensor can be calibrated to a specific target to ensure the uncertainty is as small as possible and documented if needed.

3. ESTIMATED EXPANDED UNCERTAINTY BUDGETS NCDT MEASUREMENT USING THE SIMPLIFIED METHOD

When conducting a measurement uncertainty budget in a machine tool. Machine tool users may use ISO230-2 as a basis. Employing the simplified method, it could be possible to under or overestimate the errors. By providing an example of this approach, I hope to identify the importance of understanding the application and compare it with a more detailed measurement uncertainty budget [5].

3.1. Typical measurement uncertainty for measuring thermal effects on a machine tool in industrial conditions

The below information regarding the sensor is included in the evaluation but variables used that would cause thermal effects to the measurements are also included. The table below shows an uncertainty budget using a simplified method in accordance with ISO230-2:2014 of how the thermal effects could affect the accuracy of the thermal measurement without fully understanding the application and without ensuring robust procedures and processes are in place.

Table 6: measurement uncertainty budget in accordance with ISO230-2:2014, Annex A. [5]

| Contributors | Parameter | Unit | Uncertainty | Unit |
|------------------------------------------------------------------------------------------|-----------|-----------------------------------------------|-------------|---------------|
| Difference to 20 °C, maximum 5,000 | 5 | °C | | |
| Thermal expansion coefficient 12,000 $\mu\text{m}/(\text{m} \cdot ^\circ\text{C})$ Steel | 12 | $\mu\text{m}/(\text{m} \cdot ^\circ\text{C})$ | | |
| Error range | 0.00001 | | 0.0006 | μm |
| Thermal expansion coefficient 1,500 $\mu\text{m}/(\text{m} \cdot ^\circ\text{C})$ Invar | 1.5 | $\mu\text{m}/(\text{m} \cdot ^\circ\text{C})$ | | |
| Error range | 0.3 | | 2.25 | μm |
| Device | | | | |
| Measurement length | 1 | mm | | |
| Warm up cycle | 2.111 | μm | | μm |
| Temperature sensor | 1.136 | μm | | |
| Temperature electronics | 2.121 | μm | | |
| Calibration Uncertainty | 2.5 | μm | | |
| UDEVICE | | | 7.868 | μm |
| Fixture | | | | |
| Height, assumed 200 mm | 200 | mm | | |
| Width, assumed 100 mm | 100 | mm | | |
| UF, fixture steel | | | 18 | μm |
| UF, fixture Invar | | | 2.25 | μm |
| Test bar | | | | |
| Height, assumed 250 mm | 250 | mm | | |
| UTB, test bar steel | | | 15 | μm |
| UTB, test bar invar | | | 1.875 | μm |
| UNCDD steel | | | 40.9 | μm |
| UNCDD invar | | | 14.2 | μm |

Based on the above data, the thermal measurement uncertainty for using the NCDT sensor and using ISO230-2 as a guide. The summation of the measurement uncertainties for steel would be $\pm 40.9\mu\text{m}$ ($k=2$) and for the Invar material selection of $\pm 14.2\mu\text{m}$ ($k=2$) with the divisor at $k = 2$ it is assumed that the level of confidence would be over 95% [5].

4. THERMAL EXPANDED UNCERTAINTY EVALUATION IN A PRACTICAL SITUATION

The following symbols and calculations made can be traced back to ISO/IEC Guide 98-3:2008 for understanding how the measurement uncertainty budget was created [6].

Table 7: Measurement uncertainty budget symbols:

| | |
|---------------------------------------------------------------|-----------------------------------------------------------|
| L_s : Length of reference standard at 20°C | v_i – Degrees of freedom |
| d : average of repeated indication values | WSf – Welch Satterthwaite formula |
| θ_0 : mean temperature deviation of gauges | f_i – fixture geometry effect |
| Δ : offset from mean temperature sensor of gauges | T_i – Temperature error of system component |
| $\delta\alpha$: difference in thermal expansion coefficients | MU – Total measurement uncertainty of sources |
| k_i – Divisor | $U_i(L)/\mu\text{m}$ – Contribution from i th quantity |
| c_i – Sensitivity coefficient | GUM – Guide to expression of uncertainty in a measurement |

The DS18B20 temperature sensors used had a standard uncertainty of 0.2°C and an accuracy of $\pm 0.5^{\circ}\text{C}$. Assumptions that the test was conducted at 25°C , so that the uncertainty related to correcting the accuracy of an engineering drawing to BS8888 was incorporated. The expanded thermal measurement uncertainty for table 8 would be:

$$U_{NUB} = k \times u(L) = 2.026 \times 6.274 = 12.713 \mu\text{m} \approx 12.8 \mu\text{m}$$

Therefore, the expanded thermal measurement uncertainty for the service would be $\pm 12.8 \mu\text{m}$ with a corresponding level of confidence of 95% and by using the Welch-Satterthwaite formula (WSf) the effective degrees of freedom can be evaluated to be 19. By compensating for the temperature effect on the sensor and electronics and providing traceability for the compensation, the processes to ensure traceability of the thermal expansion of the materials and ensuring a robust procedure to remove the effect of the warm up cycle a clearer, traceable and accurate uncertainty budget can be produced to improve the thermal analysis service.

Table 8: Measurement uncertainty budget for a thermal service in a practical application [6]

| Source | Uncertainty value | Units | Distribution | k_i | c_i | c_i Units | $U_i(L)/\mu\text{m}$ | v_i |
|-------------------------|-------------------|-------------------------|--------------|---------|---------|--------------------------------|----------------------|----------|
| Ls | Unknown | | Unknown | Unknown | Unknown | | 0.000 | ∞ |
| d | 2.5 | μm | t | 1 | 1 | μm | 2.500 | 16 |
| θ_0 | 0.2 | $^{\circ}\text{C}$ | n | 1 | 0 | $\mu\text{m}/^{\circ}\text{C}$ | 0.000 | ∞ |
| θ_{0x} | 0.2 | $^{\circ}\text{C}$ | n | 2 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.300 | 1 |
| θ_{0y} | 0.2 | $^{\circ}\text{C}$ | n | 2 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.300 | 1 |
| θ_{0z} | 0.2 | $^{\circ}\text{C}$ | n | 2 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.300 | 1 |
| Δ | 0.5 | $^{\circ}\text{C}$ | U | 1.414 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.000 | ∞ |
| Δx | 0.5 | $^{\circ}\text{C}$ | U | 1.414 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.750 | 1 |
| Δy | 0.5 | $^{\circ}\text{C}$ | U | 1.414 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.750 | 1 |
| Δz | 0.5 | $^{\circ}\text{C}$ | U | 1.414 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.750 | 1 |
| $\delta\alpha$ | 0.000001 | $^{\circ}\text{C}^{-1}$ | R | 1.732 | 1 | $\mu\text{m}/^{\circ}\text{C}$ | 0.000 | ∞ |
| fx | 0.2 | $^{\circ}\text{C}^{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.173 | 1 |
| fy | 0.1 | $^{\circ}\text{C}^{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.087 | 1 |
| fz | 0.1 | $^{\circ}\text{C}^{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.087 | 1 |
| tb | 0.25 | $^{\circ}\text{C}^{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^{\circ}\text{C}$ | 0.217 | 1 |
| Ts | 0.045 | μm | t | 1 | 1 | | 0.045 | 1 |
| Te | 0.016 | μm | t | 1 | 1 | | 0.016 | 1 |
| Corrections | | | | | | | | |
| $\delta\alpha$ | 0.3 | $^{\circ}\text{C}^{-1}$ | R | 1 | 7.5 | $\mu\text{m}/^{\circ}\text{C}$ | 2.250 | 1 |
| Ts | 1.136 | | t | 1 | 1 | | 1.136 | 1 |
| Te | 2.12 | | t | 1 | 1 | | 2.12 | 1 |
| Total (μm) | | | | | | | 6.274 | |
| MU (\pm) | | | | | | | 12.713 | |

The calculation for the measurement uncertainty parameters in table 9, table 11 and table 13 can be found within the GUM and the level of confidence is a function of the divisors within the specific uncertainty budget which based on the summation of the divisors, a k value of 1 would be approximately 68% and so on [3].

Table 9: Measurement uncertainty budget values calculated from the GUM for table 8 [6]

| Measurement Uncertainty budget parameter | Value |
|------------------------------------------|--------|
| k | 2.026 |
| $U_i^2(L)$ | 2.873 |
| WSf | 19.935 |

The accuracy of the measurement can be reduced further by concentrating on one axis. A new evaluation was performed with all the aforementioned process improvements included. The results are shown in table 10.

The thermal measurement uncertainty budget has been reduced by approximately 69% from $\pm 14.2\mu\text{m}$ to $\pm 4.5\mu\text{m}$.

Table 10: Measurement uncertainty budget for a robust and traceable service [6]

| Source | Uncertainty value | Units | Distribution | k_i | c_i | c_i Units | $U_i(L)/\mu\text{m}$ | v_i |
|-------------------------|-------------------|-----------------------|--------------|---------|---------|------------------------------|----------------------|----------|
| Ls | Unknown | | Unknown | Unknown | Unknown | | 0.000 | ∞ |
| d | 2.5 | μm | t | 1 | 1 | | 2.500 | 16 |
| θ_{0z} | 0.2 | $^\circ\text{C}$ | n | 2 | 1.5 | $\mu\text{m}/^\circ\text{C}$ | 0.300 | 1 |
| Δz | 0.5 | $^\circ\text{C}$ | U | 1.414 | 1.5 | $\mu\text{m}/^\circ\text{C}$ | 0.750 | 1 |
| $\delta\alpha$ | 0.000001 | $^\circ\text{C}_{-1}$ | R | 1.732 | 1 | $\mu\text{m}/^\circ\text{C}$ | 0.000 | ∞ |
| fz | 0.1 | $^\circ\text{C}_{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^\circ\text{C}$ | 0.087 | 1 |
| tb | 0.25 | $^\circ\text{C}_{-1}$ | R | 1.732 | 1.5 | $\mu\text{m}/^\circ\text{C}$ | 0.217 | 1 |
| Ts | 0.045 | μm | t | 2 | 1 | | 0.023 | 1 |
| Te | 0.016 | μm | t | 2 | 1 | | 0.008 | 1 |
| Corrections | | | | | | | | |
| $\delta\alpha$ | 0.3 | $^\circ\text{C}_{-1}$ | R | 1 | 7.5 | $\mu\text{m}/^\circ\text{C}$ | 2.250 | 1 |
| Ts | 1.136 | μm | t | 1 | 1 | | 1.136 | 1 |
| Te | 2.12 | μm | t | 1 | 1 | | 2.12 | 1 |
| Total (μm) | | | | | | | 3.884 | |
| MU (\pm) | | | | | | | 4.405 | |

Table 11: Measurement uncertainty budget values calculated from the GUM for table 10 [6]

| Measurement Uncertainty budget parameter | Value |
|------------------------------------------|--------|
| k | 1.134 |
| $U_i^2(L)$ | 2.629 |
| WSf | 17.252 |

The measurement uncertainty was reduced further by conducting numerous tests to functionally prove out the system and perform a Type A evaluation/ statistical analysis of the displacement and temperature sensor system performance. However, due to the nature of thermal tests and their long duration, this would be very difficult to achieve in practise. Also, by applying numerous measurement sensors in proximity to the system and machine tool an improved understanding of the system can be applied which in turn can improve the accuracy and the level of confidence. By using the NCDT sensor system and conditioning monitoring machine tools under the current uncertainty budget the aim is to reduce the uncertainty and improve the level of confidence to the below table.

Table 12: Condition monitoring and thermal analysis service to improve the accuracy by Type A evaluation [6]

| Source | Uncertainty value | Units | Distribution | k_i | c_i | c_i Units | $U_i(L)/\mu m$ | v_i |
|-------------------|-------------------|------------------|--------------|---------|---------|-------------------|----------------|----------|
| Ls | Unknown | | Unknown | Unknown | Unknown | | 0.000 | ∞ |
| d | 2.5 | μm | t | 6 | 1 | μm | 0.417 | 100 |
| θ_{0z} | 0.2 | $^{\circ}C$ | n | 2 | 1.5 | $\mu m/^{\circ}C$ | 0.300 | 1 |
| Δz | 0.5 | $^{\circ}C$ | t | 6 | 1.5 | $\mu m/^{\circ}C$ | 0.750 | 100 |
| $\delta\alpha$ | 0.000001 | $^{\circ}C_{-1}$ | t | 1.732 | 1 | $\mu m/^{\circ}C$ | 0.000 | ∞ |
| fz | 0.1 | $^{\circ}C_{-1}$ | t | 1.732 | 1.5 | $\mu m/^{\circ}C$ | 0.087 | 1 |
| tb | 0.25 | $^{\circ}C_{-1}$ | t | 1.732 | 1.5 | $\mu m/^{\circ}C$ | 0.217 | 1 |
| Ts | 0.045 | μm | t | 2 | 1 | | 0.023 | 1 |
| Te | 0.016 | μm | t | 2 | 1 | | 0.008 | 1 |
| Corrections | | | | | | | | |
| $\delta\alpha$ | 0.3 | $^{\circ}C_{-1}$ | R | 1 | 7.5 | $\mu m/^{\circ}C$ | 2.250 | 1 |
| Ts | 1.136 | | t | 1 | 1 | | 1.136 | 1 |
| Te | 2.12 | | t | 1 | 1 | | 2.12 | 1 |
| Total (μm) | | | | | | | 1.800 | |
| MU (\pm) | | | | | | | 3.480 | |

Table 13: Measurement Uncertainty budget values calculated using the GUM for Table 12 [6]

| Measurement Uncertainty budget parameter | Value |
|------------------------------------------|--------|
| k | 1.933 |
| $U_i^2(L)$ | 0.913 |
| WSf | 50.353 |

By controlling the service sufficiently and gathering large amounts of data, the accuracy of this could be reduced further to within $7\mu m$ shown in table 12 with well over a 90% confidence level which can be interpreted from knowing that a k value of 2 corresponds to 95% confidence level. The service is constantly being developed with the intention of reducing the thermal measurement uncertainty to within $\pm 0.5\mu m$ for application on a range high precision machine tools.

5. CONCLUSIONS

- It is important to know how the measurement uncertainty budget was created to be shown with the calibration certificate to ensure clarity and traceability and avoid increasing the uncertainty by duplication of unknown values.
- By knowing the measurement uncertainty of a sensor, tests can be performed with a system to check there isn't any other uncertainties generated.
- The measurement uncertainty of measuring the thermal effects of a machine tool can be reduced by 80% by employing a robust and traceable measurement service.
- Due to the complexity and duration of tests involved with thermally assessing a machine tool, it can be difficult for a manufacturer or machine shop to be able to ensure that the accuracy of the service is conducted. By sub-contracting the expertise out, the continuous improvement potential of this service could dramatically improve the quality of manufactured components.

REFERENCES

- [1] British Standards Institution. (2007). 230: Test code for machine tools - Part 3: Determination of thermal effects. London: BSI.
- [2] Microepsilon. (2016). *Eddy current sensors for displacement and position*. [Brochure].
- [3] Mekid, S. (2009). Introduction to Precision Machine Design and Error Assessment. Hoboken: CRC Press.
- [4] Zhang, J; Feng, P; Chen, C; Yu, D. (2013). A method for thermal performance modeling and simulation of machine tools. The international Journal of Advanced Manufacturing Technology, 1517-1527.
- [5] British Standards Institution. (2014). 230: Test code for machine tools - Part 2: Determination of accuracy and repeatability of positioning of numerically controlled axes. London: BSI.
- [6] ISO/IEC Guide 98-3:2008, Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)

Research on Spectrum Components of Planetary Gearbox Vibration Signals by Dynamic Modelling and Filtering

Lun Zhang^{1,2}, Niaoqing Hu^{1,2}, Fengshou Gu³ and Zhe Cheng^{1,2}

¹ Laboratory of Science and Technology on Integrated Logistics Support, National University of Defence Technology, Changsha Hunan, China

² College of Mechatronics and Automation, National University of Defence Technology, Changsha Hunan, China

³ Centre for Efficiency and Performance Engineering, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ABSTRACT

Understanding of planetary gearbox vibration spectrum components is helpful for researchers and engineers to develop fault diagnosis methods, however, few literatures has focus on comprehensive explanation of planetary gearbox vibration signal, there is a lack of knowledge on fully understanding of planetary gearbox vibration signal spectrum. To address this problem, a dynamic model of a planetary gearbox is established, signal of gearbox under different health state are simulated, simulation signal is processed to investigate spectrum components of planetary gear vibration signal; experiments are conducted to collect gearbox vibration signal to validate spectrum components analysis result from dynamic model simulation.

Keywords: Spectrum Components; Dynamic Modeling; Gearbox; Modeling analysis and optimization

Corresponding author: Niaoqing Hu (hnq@nudt.edu.cn)

1. INTRODUCTION

Planetary gear posses several unique advantages over parallel axis gear trains, it can provide a considerable reduction ratio in a compact space while keeping input power and output power coaxial. Application of planetary gear can be found in lots of industrial areas like aerospace, helicopter main gearbox and wind turbine. Working under heavy loads, gears in planetary gear system may suffer from various damages such as pitting^[1] and wearing^[2]. Damaged planetary gear will degrade performance of equipment by increasing vibration and noise levels, some severe damage will result in breaking down of whole power train, leading to economic losses.

Condition monitoring and fault diagnosis techniques have been introduced to ensure availability and safety of planetary gear. Researchers in this field conducted series studies to investigate effectiveness of different diagnosis methods^{[3][4]}. Most commonly seen techniques used in condition monitoring of planetary gear includes vibration analysis^[5], oil debris analysis^[6], and acoustic emission analysis^[7]. Oil debris analysis methods collect metal particles that floating in the lubricants, evaluating equipment health status according to quantity or increasing rate of metal particles, but it is hard to conduct online monitoring with oil debris analysis, what's more, once the lubricants are replaced, which is a very common thing, oil debris analysis will fail. Acoustic emission (AE) was originally developed for non-destructive testing of static structures, however, in recent years its application has been extended to health monitoring of rotating machines; but AE sensors should be close to the source, moreover, rapid stress-releasing events of interests generates a spectrum starting at 0 Hz and falling off at several MHz, thus a high sampling rate is required when using AE based methods. Vibration analysis has been approved to be an effective approach for analyzing, detecting and diagnosing gear faults in industrial systems^{[8][9][10]}, it can be easily applied to online monitoring, sampling rate of vibration signal is typically lower than 50 kHz. Vibration analysis then becomes most popular techniques in the field of diagnosis of planetary gearbox.

In vibration signal analysis, fault characteristic frequencies are vital parameters^[3]. Spectrum based analysis methods aim at finding existence or apparent increasing at given frequency components that highly relates to condition of gearbox. Time-frequency domain analysis methods aim at tracking time-varying frequency caused by varying working condition with a considerable resolution in both time and frequency domain. Besides signal processing methods, several enhancement detection methods, stochastic resonance^[11] for example, were proposed to solve the proverbial problem that fault related signals are usually weak and may be flooded by regular vibrations. Researches on planetary gearbox have unveiled dynamic modeling and fault characteristic frequency calculation formulas for different kind of dynamic components, however, few literature has focus on comprehensive explanation of planetary gearbox vibration signal, there is a lack of knowledge on fully understanding of planetary gearbox vibration signal spectrum.

To address this problem, a dynamic model of a planetary gearbox is established, signal of gearbox under health condition is firstly analyzed to investigate regular vibration components, then fault on each gear is simulated with the dynamic model, simulation signal is processed to investigate changes of regular vibration components and emerging frequency, experiments are conducted to collect gearbox vibration signal, research in this paper is validated by investing residual spectrum after filtering out known frequency with high quality factor band stop filters.

2. MODELING OF PLANETARY GEARBOX

Most widely used planetary gearbox set consists of a sun gear, a fixed ring gear and a number of identical planet gears (typically 3-6) meshing with both ring gear and sun gear, a carrier is used to hold the planet gears in place and outputs movements and power.

Shown in figure 1(a) is schematic diagram of planetary gear as well as some dynamic parameters, Parameters of each part are distinguished with subscript; carrier is represented by c , ring gear is represented by r , sun gear is represented by s and planet gear is represented by p , when emphasis number of planet gear, subscript will be $n = 1, 2, \dots, N$, where N is number of planet gear in planetary gearbox.

Without losing generality, following assumption is applied during modeling of planetary gear:

1. Every part in planetary gearbox moves in a plane that is vertical to rotating axis.
2. Planetary gearbox is simplified as a lumped parameter model, supporting and gears are simplified as spring and damping, gear body and shafts are regarded as rigid body.
3. Planet gears are evenly distributed around sun gear; parameters of each planet gear are identical.
4. For planetary gear under health condition, meshing stiffness varies as a rectangular wave.

In figure 1(a), there are several coordinates which are used to describe motion of moving parts conveniently, OXY coordinate is fixed while Oxy coordinate moves as carrier, which means Oxy is fixed with carriers, and rotating speed of Oxy is identical with angular velocity of carrier ω , x axis of Oxy passes through center of first planet gear. Coordinate $O_nx_ny_n$ also rotate as carrier, their origin are located at center of each planet gear respectively. Parameters in figure 1 are explained as follows:

x_i, y_i ($i=s, r, c, 1, 2, \dots, N$). Translational displacement of theoretical center of each moving parts; they are measured in coordinate Oxy .

u_i ($i=s, r, c, 1, 2, \dots, N$). Rotational displacement of theoretical center of each moving parts, also measured in coordinate Oxy . $u_i = r_i \theta_i$, where r_i is base circle radius for gear components or center circle of carrier, θ_i is angular displacement.

φ_n ($n = 1, 2, \dots, N$). Angle between OO_n and x axis of Oxy , $\varphi_n = 2\pi (n-1)/N$.

k_p, c_p . Supporting stiffness and damping of planet gear.

k_i, c_i ($i = c, r, s$). Radial supporting stiffness and damping of central components.

k_{it}, c_{it} ($i = c, r, s$). Tangential supporting stiffness and damping of central components.

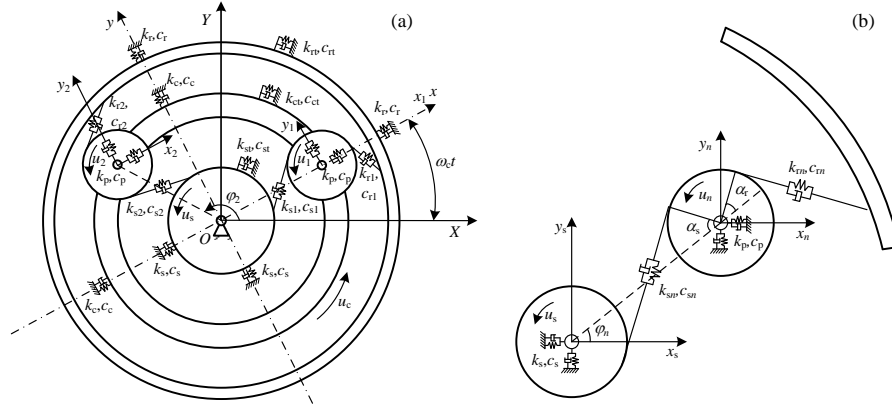


Figure 1. Planetary gearbox. (a) dynamic model (b) relative displacement.

2.1. Acceleration analysis

Acceleration analysis of each part is necessary to build a dynamic model. As stated earlier, coordinates of each moving parts are fixed with carrier, however, mass centre acceleration used in dynamic analysis must be absolute acceleration. Following are how to describe absolute acceleration of parts in moving coordinate.

Giving vector r and it is (x, y) in moving coordinate Oxy while (X, Y) in absolute coordinate, relationship between (x, y) and (X, Y) will be

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \omega_c t & \sin \omega_c t \\ -\sin \omega_c t & \cos \omega_c t \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (6)$$

First order derivative of function (6) is

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \cos \omega_c t & \sin \omega_c t \\ -\sin \omega_c t & \cos \omega_c t \end{bmatrix} \begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} + \omega_c \begin{bmatrix} -\sin \omega_c t & \cos \omega_c t \\ -\cos \omega_c t & -\sin \omega_c t \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (7)$$

Second order derivative of function (6) is

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} \cos \omega_c t & \sin \omega_c t \\ -\sin \omega_c t & \cos \omega_c t \end{bmatrix} \begin{bmatrix} \ddot{X} \\ \ddot{Y} \end{bmatrix} + 2\omega_c \begin{bmatrix} -\sin \omega_c t & \cos \omega_c t \\ -\cos \omega_c t & \sin \omega_c t \end{bmatrix} \begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} + \omega_c^2 \begin{bmatrix} -\cos \omega_c t & -\sin \omega_c t \\ \sin \omega_c t & -\cos \omega_c t \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (8)$$

According to function(7) and function(8), absolute coordinate component of absolute acceleration in moving coordinate Oxy will be

$$\begin{cases} a_x = \ddot{x} - 2\omega_c \dot{y} - \omega_c^2 x \\ a_y = \ddot{y} + 2\omega_c \dot{x} - \omega_c^2 y \end{cases} \quad (9)$$

Thus, acceleration of each part will be

$$\begin{cases} a_{ix} = \ddot{x}_i - 2\omega_c \dot{y}_i - \omega_c^2 x_i \\ a_{iy} = \ddot{y}_i + 2\omega_c \dot{x}_i - \omega_c^2 y_i \end{cases} \quad (i = s, r, c, 1, 2, \dots, N) \quad (10)$$

2.2. Relative displacement analysis

Figure 1(b) shows relative position of gears that engaged in meshing, a detailed deduction of relative displacement of sun gear and planet gear will be given in the following; other relative displacement could be deducted similarly.

Along the meshing line, direction from sun gear to planet gear is regarded as positive direction, relative displacement of sun gear and planet gear can be obtained by projecting displacement of sun gear and planet gear to meshing line.

Translational displacements of sun gear are x_s and y_s , their projection on meshing line are $-x_s \sin \varphi_{sn}$ and $y_s \cos \varphi_{sn}$, where $\varphi_{sn} = \varphi_n - \alpha_s$; translational displacements of planet gear are x_n and y_n , and their projection are $-x_n \sin \varphi_{sn}$ and $y_n \cos \varphi_{sn}$; rotational displacement of sun gear and planet gear are u_s , u_n , their projection on meshing line are u_s and u_n ; meshing error of planet gear and sun gear is $e_{sn}(t)^{[12]}$. Thus, relative displacement of sun gear and planet gear is

$$\delta_{sn} = (x_n - x_s) \sin \varphi_{sn} + (y_n - y_s) \cos \varphi_{sn} + u_s + u_n + e_{sn}(t) \quad (11)$$

2.3. Dynamic model of planetary gear

Assuming sun gear is input component while carrier is output, and input torque and output torque are T_s and T_c respectively, mass of carrier, ring gear, sun gear and planet gear are m_c , m_r , m_s and m_p , moment of inertia are I_c , I_r , I_s and I_p . According to Newton's second law, dynamic functions of each component can be obtained as follows.

Dynamic functions of carrier are

$$\left. \begin{aligned} m_c(\ddot{x}_c - 2\omega_c \dot{y}_c - \omega_c^2 x_c) + \sum_{n=1}^N (k_p \delta_{cnx} + c_p \dot{\delta}_{cnx}) + k_c x_c + c_c \dot{x}_c &= 0 \\ m_c(\ddot{y}_c + 2\omega_c \dot{x}_c - \omega_c^2 y_c) + \sum_{n=1}^N (k_p \delta_{cny} + c_p \dot{\delta}_{cny}) + k_c y_c + c_c \dot{y}_c &= 0 \\ (I_c / r_c^2) \ddot{u}_c + \sum_{n=1}^N (k_p \delta_{cnu} + c_p \dot{\delta}_{cnu}) + k_{ct} u_c + c_{ct} \dot{u}_c &= -T_c / r_c \end{aligned} \right\} \quad (12)$$

dynamic functions of ring gear are

$$\left. \begin{aligned} m_r(\ddot{x}_r - 2\omega_r \dot{y}_r - \omega_r^2 x_r) - \sum_{n=1}^N (k_{rn} \delta_{rn} \sin \varphi_{rn} + c_{rn} \dot{\delta}_{rn} \sin \varphi_{rn}) + k_r x_r + c_r \dot{x}_r &= 0 \\ m_r(\ddot{y}_r + 2\omega_r \dot{x}_r - \omega_r^2 y_r) + \sum_{n=1}^N (k_{rn} \delta_{rn} \cos \varphi_{rn} + c_{rn} \dot{\delta}_{rn} \cos \varphi_{rn}) + k_r y_r + c_r \dot{y}_r &= 0 \\ (I_r / r_r^2) \ddot{u}_r + \sum_{n=1}^N (k_{rn} \delta_{rn} + c_{rn} \dot{\delta}_{rn}) + k_{rt} u_r + c_{rt} \dot{u}_r &= 0 \end{aligned} \right\} \quad (13)$$

dynamic functions of sun gear are

$$\left. \begin{aligned} m_s(\ddot{x}_s - 2\omega_s \dot{y}_s - \omega_s^2 x_s) - \sum_{n=1}^N (k_{sn} \delta_{sn} \sin \varphi_{sn} + c_{sn} \dot{\delta}_{sn} \sin \varphi_{sn}) + k_s x_s + c_s \dot{x}_s &= 0 \\ m_s(\ddot{y}_s + 2\omega_s \dot{x}_s - \omega_s^2 y_s) + \sum_{n=1}^N (k_{sn} \delta_{sn} \cos \varphi_{sn} + c_{sn} \dot{\delta}_{sn} \cos \varphi_{sn}) + k_s y_s + c_s \dot{y}_s &= 0 \\ (I_s / r_s^2) \ddot{u}_s + \sum_{n=1}^N (k_{sn} \delta_{sn} + c_{sn} \dot{\delta}_{sn}) + k_{st} u_s + c_{st} \dot{u}_s &= T_s / r_s \end{aligned} \right\} \quad (14)$$

dynamic functions of planet gear n are

$$\left. \begin{aligned} m_p(\ddot{x}_n - 2\omega_c \dot{y}_n - \omega_c^2 x_n) + k_{sn} \delta_{sn} \sin \varphi_{sn} + c_{sn} \dot{\delta}_{sn} \sin \varphi_{sn} + k_{tn} \delta_{tn} \sin \varphi_{tn} + c_{tn} \dot{\delta}_{tn} \sin \varphi_{tn} - k_p \delta_{cny} - c_p \dot{\delta}_{cny} &= 0 \\ m_p(\ddot{y}_n - 2\omega_c \dot{x}_n - \omega_c^2 y_n) - k_{sn} \delta_{sn} \cos \varphi_{sn} - c_{sn} \dot{\delta}_{sn} \cos \varphi_{sn} - k_{tn} \delta_{tn} \cos \varphi_{tn} - c_{tn} \dot{\delta}_{tn} \cos \varphi_{tn} - k_p \delta_{cny} - c_p \dot{\delta}_{cny} &= 0 \\ (I_p / r_p^2) \ddot{u}_s + k_{sn} \delta_{sn} + c_{sn} \dot{\delta}_{sn} - k_{tn} \delta_{tn} - c_{tn} \dot{\delta}_{tn} &= 0 \end{aligned} \right\} \quad (15)$$

where k_{sn} and k_{tn} are time-varying stiffness of external meshing and internal meshing respectively; δ_{sn} , δ_{tn} , δ_{cny} , δ_{cny} are relative displacements.

3. SPECTRUM COMPONENTS ANALYSIS BASED ON SIMULATION

Taking advantage of dynamic model built above, dynamic model of test rig planetary gear is established; parameters of each gear are measured and listed in table 1.

Table 1: parameters of gear components

| Component | Tooth Number | Pitch Circle Diameter(mm) | Mass(kg) |
|-------------|--------------|---------------------------|----------|
| Ring gear | 96 | 192 | 2.62 |
| Sun gear | 28 | 56 | 0.38 |
| Planet gear | 34 | 68 | 0.44 |
| Carrier | -- | 124 | 3.8 |

During simulation, carrier rotating speed is set to 240 rpm; fault characteristic frequencies are calculated and listed in table 2. Formulas to calculate characteristic frequency could be found in [3].

Table 2: Fault characteristic frequency of simulation signal

| Meshing Frequency | Rotating Frequency | | Distributed Faults | | | Local Faults | | |
|-------------------|--------------------|---------|--------------------|---------|---------|--------------|---------|---------|
| | sun | carrier | ring | planet | sun | ring | planet | sun |
| 384Hz | 17.71Hz | 4Hz | 4Hz | 11.29Hz | 13.71Hz | 16Hz | 11.29Hz | 54.84Hz |

Planetary gear under health condition is first simulated as baseline test, then planetary gear with different faults is simulated as well, spectrum of simulation signal is shown in figure 4.

Obviously, under health condition, spectrum of planetary vibration signal is dominated by meshing frequency; vibration signal is modulated by rotating of carrier, which introduced sidebands around meshing frequency in figure 4(a). when a gear component is damaged in planetary gearbox, taking planet gear as an example, more sidebands will appear around meshing frequency besides carrier rotating frequency, a comparison of figure 4(c) and table 2 will indicates that the spacing between sidebands is fault characteristic frequency. In figure 4(c), some tiny spectrum components are also noticeable besides fault characteristic frequency, a zoom-in view of figure 4(c) is shown in figure 5, second and third order sidebands(namely 356.58 and 342.87) and tiny frequency components are shown, obviously, the spacing between tiny frequency components and sidebands is identical to rotating frequency of carrier, this phenomenon is unique character of planetary gearbox since sun gear fault related vibrations could not be measured directly, sensors around ring gear amount on the cage have to 'feel' sun gear vibration through planet gear that travels along the carrier.

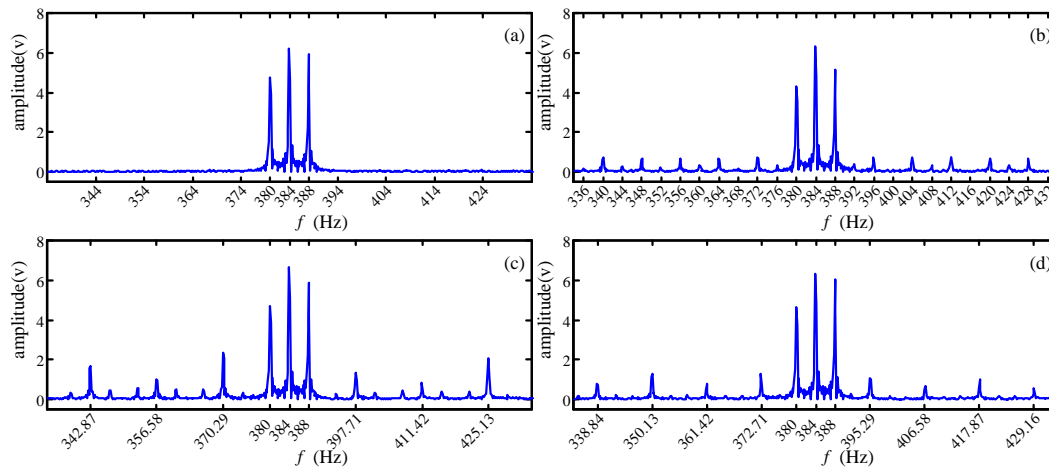


Figure 2. Spectrum of simulation signal. (a) health (b) ring gear fault (c) sun gear fault (d) planet gear fault.

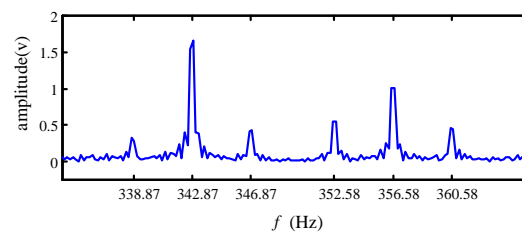


Figure 3. Zoon-in view of figure 4(c).

Spectrum components analysis of simulation signal indicates that spectrum of planetary gear is dominated by gear meshing frequency and modulation of carrier rotating frequency, when a gear is damaged, corresponding fault characteristic frequency will appear in the spectrum, if sun gear is damaged, there will also be sidebands near fault characteristic sidebands with a spacing that equals to carrier rotating frequency.

4. EXPERIMENT SIGNAL VALIDATION

To validate spectrum component analysis result in section 3, experiments are conducted on a planetary test rig. During the experiments, vibration signals are collected and analyzed afterward.

Figure 4 shows the planetary gear test rig and its schematic figure. Tests are conducted on second stage planetary gear. Parameters of each part are measured and listed in table 1 (in section 3). Running speed of drive motor is 1200 rpm and load is 41.2 Nm.

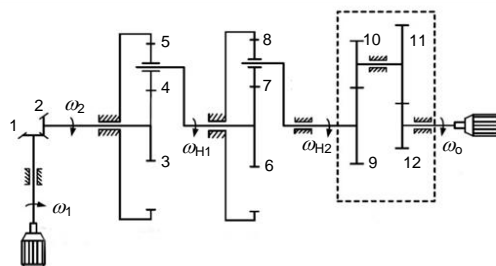


Figure 4. Planetary gearbox test rig.

Two accelerometers are mounted on second planetary gear case to collect vibration data on horizontal and vertical direction synchronously. Signals are collected by a PC-based DAQ system after filtered by an anti-

aliasing filter. The limit frequency is 2 kHz and sampling rate is 5.12 kHz, 198.4s data are collected during each test. Fault characteristic frequencies are listed in table 3.

Table 3: Fault characteristic frequency of experiment signal

| Meshing Frequency | Rotating Frequency | | Distributed Faults | | | Local Faults | | |
|-------------------|--------------------|----------|--------------------|---------|--------|--------------|---------|--------|
| | sun | carrier | ring | planet | sun | ring | planet | sun |
| 48.16Hz | 2.221Hz | 0.5016Hz | 0.5016Hz | 1.416Hz | 1.72Hz | 2.0Hz | 1.416Hz | 6.88Hz |

A series of band stop filters are used to validate spectrum component analysis result in section 3. For the purpose of filtering out specific frequency component and in case of filtering out other frequency components unconsciously, high quality factors are chosen for band stop filters. Taking meshing frequency as an example, band stop filter centre frequency is 48.16Hz, stop band width must be smaller than carrier rotating frequency, otherwise, carrier rotating sidebands will be influenced when filtering out meshing frequency, thus quality factor should be larger than 96.

Validation is conducted as follows: first, meshing frequency is filtered out, then carrier rotating frequency sidebands are filtered, after that fault characteristic frequency of sun gear and sidebands around then are filtered; finally, fault characteristic frequency of planet gear and ring gear are filtered respectively. If the analysis result in section 3 is correct, there will be no apparently dominate frequency component in the spectrum.

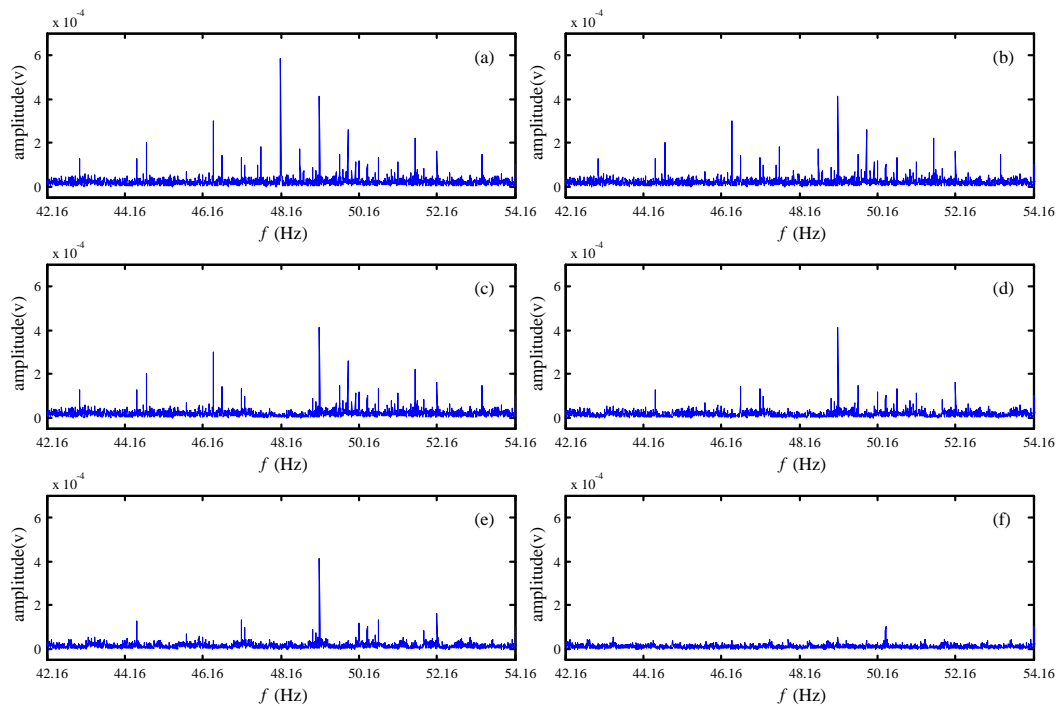


Figure 5. Spectrum components of experimental signal (a) original spectrum (b) meshing frequency filtered (c) carrier rotating frequency sidebands filtered (d) sun gear fault sidebands filtered (e) planet gear fault sidebands filtered (f) ring gear fault sidebands filtered

Figure 5 shows spectrum of planetary gear experimental signal, figure 5(a) is spectrum of original signal, (b),(c), (d),(e), (f) are residual spectrum after each filtering step. Obviously, after filtering spectrum components included in analysis results, no apparently dominate frequency component in the spectrum is left. Experimental signal validation indicates that spectrum components analysis by dynamic modelling is effective.

5. CONCLUSION

Understanding of planetary gearbox vibration spectrum components is helpful for researchers and engineers in this field to develop fault diagnosis methods, based on dynamic modelling and filtering, spectrum components of planetary gearbox vibration signal under different health condition are analyzed by filtering out known frequency with a series of high quality factor band stop filters; results are validated by experimental signal collected from a planetary test rig.

ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (51475463, 51575518). Thanks also go to the anonymous reviewers and the editor for their suggestions and comments, which have helped improving the quality of this paper.

REFERENCES

- [1] Xu, X., Dong, S., Liao, C., Tao, Y., & He, Z. (2016). Research on dynamical characteristics of planetary gear system with tooth pitting. *Journal of Vibroengineering*, 18(7).
- [2] Tian, H., Kang, X. Y., Li, Y. J., & Zhang, J. N. (2014). Fault Diagnosis of Gear Wearing Based on Order Cepstrum Analysis. In *Applied Mechanics and Materials* (Vol. 543, pp. 922-925). Trans Tech Publications.
- [3] Lei, Y., Lin, J., Zuo, M. J., & He, Z. (2014). Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement*, 48, 292-305.
- [4] Park, J., Ha, J. M., Oh, H., Youn, B. D., Choi, J. H., & Kim, N. H. (2016). Model-Based Fault Diagnosis of a Planetary Gear: A Novel Approach Using Transmission Error. *IEEE Transactions on Reliability*, 65(4), 1830-1841.
- [5] Liang, X., Zuo, M. J., & Hoseini, M. R. (2015). Vibration signal modeling of a planetary gear set for tooth crack detection. *Engineering Failure Analysis*, 48, 185-200.
- [6] Shah, H., & Hirani, H. (2014). Online condition monitoring of spur gears. *International Journal of Condition Monitoring*, 4(1), 15-22.
- [7] Vicuña, C. M. (2014). Effects of operating conditions on the Acoustic Emissions (AE) from planetary gearboxes. *Applied Acoustics*, 77, 150-158.
- [8] Z. Man, W. Wang, S. Khoo, J. Yin, Optimal sinusoidal modeling of gear mesh vibration signals for gear diagnosis and prognosis, *Mechanical Systems and Signal Processing* 33 (2012) 256–274.
- [9] W. Wang, Early detection of gear tooth cracking using the resonance demodulation technique, *Mechanical Systems and Signal Processing* 15 (2001) 887–903.
- [10] R. Isermann, Model-based fault-detection and diagnosis-status and applications, *Annual Reviews in Control* 29 (2005) 71–85.
- [11] Lei, Y., Han, D., Lin, J., & He, Z. (2013). Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. *Mechanical Systems and Signal Processing*, 38(1), 113-124.
- [12] Inalpolat, M., & Kahraman, A. (2010). A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors. *Journal of Sound and Vibration*, 329(4), 371-393.

Session 8

Sensor Technology and Damage Detection 2

Session Chair Dr. Mahmood Shafiee

Combining Model-based Monitoring and a Physics of Failure Approach for Wind Turbine Failure Detection (*J. Tautz-Weinert, S. J. Watson*)

On the Feasibility of Inferring the Applied Mechanical Loading of a Conveyor System Test Rig From Monitored System Parameters (*O. Gebler, B. Hicks, A. Harrison, M. Barker*)

Assessing Uneven Milling Cutting Tool Wear using Component Measurement (*Z. J. Ahmed, P. W. Prickett, R. I. Grosvenor*)

Potential Use Cases of LWIR Cameras for Automation and Process Monitoring (*B. Eichentopf, R. Baltes, N. Fietz, K. Nienhaus*)

Combining Model-Based Monitoring and a Physics of Failure Approach for Wind Turbine Failure Detection

J. Tautz-Weinert¹, S. J. Watson¹

¹ Centre for Renewable Energy Systems Technology, Loughborough University, Loughborough LE11 3TU, UK

ABSTRACT

Condition monitoring of wind turbines with only operational data has received more attention in the last decade due to the advantage of freely available data without extra equipment needed. Although the operational data recorded by the Supervisory Control And Data Acquisition (SCADA) system are intended for performance monitoring and typically stored only every 10 minutes, information on the turbine's health can be extracted. A major focus is here on the temperature signals of mechanical parts such as drivetrain bearings. Despite the fact that absolute temperatures rise very late in the case of a failure, the temperature behaviour might change well in advance. Model-based monitoring is a tool to detect these small changes in the temperature signal affected by varying load and operation. Data-driven models are trained in a period where the turbine can be assumed to be healthy and represent the normal operation thereafter. Degradation and imminent failures can be detected by analysing the residual of modelled and measured temperatures. However, detecting failures in the residual is not always straightforward due to possibly unrepresentative training data and limited capabilities of this approach. A different way of using SCADA data lies in the estimation of damage accumulation with performance parameters based on the Physics of Failure methodology. A combination of model-based monitoring with damage calculation based on a Physics of Failure approach is proposed to strengthen the failure detection capabilities. The monitoring performance is evaluated in a case study with SCADA data from a wind farm.

Keywords: wind turbines, SCADA, physics-of-failure, condition monitoring, machine learning.

Corresponding author: J. Tautz-Weinert (email: j.tautz-weinert@lboro.ac.uk)

1. INTRODUCTION

With the exponential growth of wind energy in the last decades, the demand for optimised asset management of wind turbines has slowly evolved. In the early days of wind energy, scheduled and corrective maintenance were the appropriate measures for easy-to-access onshore farms and small turbines. With the move offshore and turbine capacities in the multi-MW category in recent years, the more complicated accessibility and significant financial losses for any downtime demand an optimised maintenance strategy. Condition-based or predictive maintenance as a proven strategy in other industries, promises to increase the efficiency of maintenance by optimising the point of intervention based on the condition of the system and risks of imminent failures.

Condition-based maintenance requires adequate measurements and monitoring techniques to reveal the health of the turbine and probabilities of upcoming failures. Due to the complexity of a wind turbine, a single measurement cannot cover the monitoring of all possible structural, mechanical and electrical failures. Failure analyses showed that the gearbox and generator are the most critical subassemblies in terms of failure rate and the corresponding downtime [1,2]. Accordingly, research and industry have focused on condition monitoring of the underlying mechanical failure mechanisms in wind turbine drive trains, although structural health monitoring of the blades, tower and foundation and detection of faults in the power converter, pitch and yaw systems have also been investigated. Potential measurements were found as vibration, acoustic emission, strain, torque, temperatures or oil parameters combined with signal processing techniques such as filtering, synchronous sampling, Hilbert transform, Wavelet transform, Fast Fourier Transform and many others [3,4].

More recently, the use of the operational data recorded by the Supervisory Control And Data Acquisition (SCADA) system has been investigated due the availability of such data without additional sensor installations. While these data are mainly intended for monitoring the performance of turbines in terms of power production, availability, possible misalignment and similar, several different applications to condition monitoring have been identified. Alarm logs in SCADA data might be analysed to find the root causes of events [5,6]. However, the most promising information for drive train condition monitoring lies in the temperature signals as mechanical degradation shows in increased thermal losses [7]. Drive train temperatures in wind turbines fluctuate with changing wind speed, rotational speed and loading. Accordingly, absolute temperature thresholds are known to give late alarms in contrast to vibration-based condition monitoring systems [3]. To overcome this drawback, model-based monitoring can reveal hidden trends in the temperature time series. Due to the complexity of wind turbine systems, data-driven learning is preferred to analytical building of models. Inputs for modelling drive train temperatures might be other temperatures, control signals as the power output or rotational speed or even the history of the target in a partly autoregressive approach. Modelling of the temperatures has been investigated with simple linear sums of inputs [8], artificial neural networks (ANNs) [9,10], adaptive neuro-fuzzy inference systems [11] or state estimation techniques [12]. A previous comparative study of the authors showed that most of the (non-autoregressive) techniques result in similar accurate prediction with slight advantages of ANNs [13].

In contrast to the model-based monitoring investigating temperature signals, the Physics of Failure approach tries to analyse the operational statistics derived from SCADA data in order to estimate the damage accumulation. In a case study with a big farm, it has been demonstrated that turbines with gearbox problems might be identified by their operational statistics [14].

In this paper, a combination of model-based monitoring with statistical analyses as used in the Physics of Failure approach is discussed and tested in a case study with data from an onshore wind farm.

2. MONITORING WIND TURBINE DRIVE TRAINS WITH OPERATIONAL DATA

The SCADA system in wind turbines usually measures multiple parameters with a sampling frequency of 1 Hz. Due to the fact that these measurements are originally intended for long-time performance monitoring, usually only averages and possibly extrema and standard deviations of ten minutes are recorded. The number and selection of measured signals depends on the turbine manufacturer or SCADA system provider, but wind speed and direction, pitch and yaw angles, rotational speed, power output and ambient temperature are always monitored. Additionally, temperatures of parts in the drive train are often measured – although with different levels of detail, e.g. only a generator and a gearbox temperature in one setup or more than twenty temperatures at different locations at the shaft in a more detailed configuration. The numerical SCADA data are supplemented by the alarm log listing all fault events happening during the operation.

2.1. Normal behaviour modelling of SCADA temperatures

Model-based monitoring [8–13] tries to identify anomalies in a system by comparing measured parameters with outputs of a model of the system. This kind of monitoring is able to highlight slight changes in measured signals affected by complex interaction of loading and heat transfers as in the wind turbine drive train. The model needs to predict the fluctuations of the temperature accurately enough to allow the residual of measured and modelled temperature to act as an indicator for possible degradation and imminent failure, as sketched in figure 1.

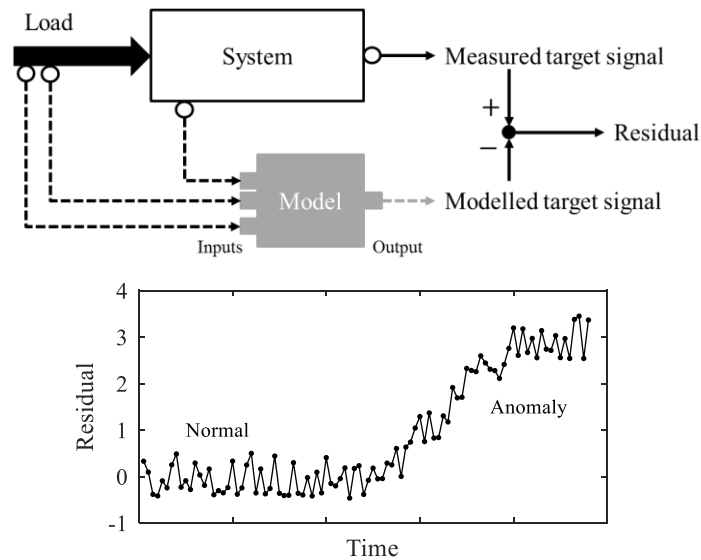


Figure 1. Sketch of model-based monitoring and indication of anomalies in the residual.

Although the basic heat generation in the drivetrain can be traced back to mechanical losses proportional to the acting wind and the rotational speed, the system is affected by more complex interaction of sub-systems, the ambient temperature and cumulative effects which make analytical modelling difficult. In contrast, data-driven modelling requires only a representative training period to learn the relationship. During this training phase the system needs to be in normal condition to enable detecting anomalies thereafter based on the difference to this behaviour. ANNs are a tool to learn and represent non-linear relationships inspired by the human brain. A common feedforward ANN trained by Levenberg-Marquardt backpropagation consists of one input layer, one or more hidden layers with a specified number of neurons and the output layer. Each neuron sums the weighted outputs of the previous layers and uses a non-linear activation function, typically a hyperbolic tangent, to generate an output. For the application of modelling a drivetrain temperature, a single linear output is used.

The inputs for modelling can be chosen based on the understanding of the system (also called domain knowledge) or based on the properties of the signals, e.g. the correlation of signals. Although using partly autoregressive modelling might increase the accuracy of prediction, this will not necessarily improve the anomaly detection capability as the prediction is influenced by the target signal and could adapt to changes in the behaviour.

Wind turbine drivetrains usually consist of main bearings, main shaft, a gearbox build of a planetary and two parallel stages, the generator shaft and generator and multiple bearings. All possible target temperatures have to be monitored as behavioural changes might not only show up in the nearest sensor, but also in other signals.

Any significant maintenance or replacement will alter the behaviour of the system. Accordingly, normal behaviour models need to be re-trained after such events.

The model-based monitoring of drivetrain temperatures aims to detect slow degradation due to mechanical wear in bearings and gears. Early identification of these problems will enable the operator to optimise the maintenance scheduling and prevent long downtimes. However, challenges in representative training and limited detection capabilities result in significant uncertainties of this monitoring approach.

2.2. Physics of Failure

The Physics of Failure approach [14] aims to estimate damage accumulation based on a simplified physical model and operational statistics derived from SCADA data. Maintenance is to be targeted based on probabilities of failures. The basis of a Physics of Failure approach is a system analysis which includes a

detailed system definition, potential failure modes with their causes and damage driving operating conditions. A damage accumulation model has to be built for each of the identified potential failure modes. Gray and Watson [14] gathered failure root causes of wind turbine gearboxes and derived several performance parameters from SCADA data to identify failure modes in a case study. The farm-wide comparison of the parameters such as average wind speed, rated power hours, brake application count, yaw movement, low speed and high power and rated speed hours, rotor starts and power dynamic, indicated that the failing turbines were affected by ‘high cycle fatigue due to poor contact between roller and raceway occurring at conditions of high stationary power’ [14]. A bearing damage model based on Lundberg-Palmgren’s bearing life formulae and linear Palmgren-Miner damage accumulation was proposed and applied using the SCADA signals power and rotational speed to approximate the bearing load. The damage model was only calibrated with the observed failures, but the resulting damage values of the failing turbines were clearly higher than the 75% percentile of the farm. However, in terms of indicating problems in certain turbines, the farm-wide comparison of the rated power hours gave similarly helpful information. Accordingly, evaluating performance parameters can be prioritised over developing full damage models.

3. CASE STUDY

In this study, data from 12 turbines in an onshore UK wind farm with a capacity of approx. 1-3 MW are analysed. The SCADA records are available from a period of 2.5 years and consist of signals in 10 minute resolution as listed in table 1, available as averages (mean) and partly maximums (max), minimums (min) and standard deviations (std). No detailed specification of sensor types or locations is available. The temperature signals are numbered, but lack a descriptive labelling.

The investigated turbines were affected by several drivetrain subassembly or part replacements, which are gathered from a commented stoppage list as the only maintenance documentation. Five gearbox replacements, three generator replacements and six bearing replacements took place. Sufficient details to describe the failure are only given for one gearbox replacement, where gear teeth broke on the intermediate speed stage gear. Only three of the investigated turbines did not undergo any major replacement.

Table 1: Case study SCADA signals

| Parameter | Signal |
|----------------------------------------------------------------------|---------------------|
| Wind speed | Mean, max, min, std |
| Wind, nacelle and relative direction | Mean |
| Pitch angle | Mean |
| Generator speed | Mean, max, min, std |
| Electrical power | Mean, max, min, std |
| Power factor, frequency | Mean |
| Voltage and current per phase | Mean |
| 16 temperatures | Mean |
| Active time for line, turbine, wind, ambient temperature, yaw motion | Seconds of 600 |

Due to the missing temperature labels in this case study, the different failing parts cannot be targeted directly by normal behaviour modelling. Instead, all temperature signals are analysed and possibly helpful targets identified. Pre-processing is applied in terms of a validity check and removal of a complete sample if invalid values are found. ANN models with 20 neurons in one hidden layer are trained with data representing 3 months. Five inputs are automatically selected on the basis of the strongest correlation in the training phase. Re-training of models after major replacements or obvious system modifications is implemented. Residuals are filtered for steps $> 5^{\circ}\text{C}$ in the target, model prediction or residual. To reduce the fluctuations, residuals are smoothed by calculating the median of each 288 samples (two days). Warnings are generated based on a threshold representing 2% exceeding probability derived from a fitted Gaussian distribution to the residual from the training period. Alarms are raised only if more than 3 of possibly 10 warnings occur in a moving window.

As a first step of the Physics of Failure approach, performance parameters are defined as given in table 2. Due to the distribution of replacements in time, analysing statistics of the whole data as done in [14] would not be helpful. In contrast, the parameters are calculated for each month accumulating all data up to this date. Adequate normalisation is chosen to enable comparing of parameters from different data size. It has to be noted that the small number of turbines in this case study impedes any statistical analysis.

Table 2: Definition of performance parameters for failure analysis. All parameters (except TUS) are calculated for operation only by requiring power mean $> 10\%$.

| Parameter | Definition | Normalisation / scaling |
|--------------------------------|----------------------------------------------|--------------------------------|
| Wind speed (WS) | Average of wind speed mean | 1.0 to 1.5 rated wind speed |
| Turbulence (TU) | Average of wind speed std | 0 to 1.5 rated wind speed * 10 |
| Turbulence in standstill (TUS) | Average of wind speed std (power $< 10\%$) | 0 to rated wind speed * 10 |
| Rated power (RP) | Count if power mean $> 90\%$ | Ratio: divide by sample size |
| High wind speed (HW) | Count if wind speed max $>$ rated wind speed | Ratio: divide by sample size |
| Power factor inverse (PF) | $1 - \text{average of power factor mean}$ | *100 |
| Power dynamic (PD) | Average of power std | 0 to rated power * 10 |
| High rotational speed (HS) | Count if generator speed mean $> 90\%$ | Ratio: divide by sample size |

4. RESULTS

4.1. Model-based monitoring

Two temperatures are identified to relate to gearbox failures. The advance detection of problems is demonstrated in figures 2 and 3. Gearbox problems are detected 39, 66, 75, 78 and possibly 492 days in advance for the five gearbox replacements, respectively. However, if the approach is applied to all turbines, a significant number of alarms is issued without known gearbox problems, see figure 4. The alarms might be false or indicate other unreported problems. If the generator failures are to be detected, using another temperature shows good indication for the two replacements in the same turbine. However, the number of alarms in other turbines without generator replacement is high, see figure 5. The alarm distribution over time indicates here a seasonal pattern visible in most turbines. Additionally, it seems possible, that some alarms might indicate gearbox problems. No clear alarm pattern is found in any of the temperatures for the bearing replacements.

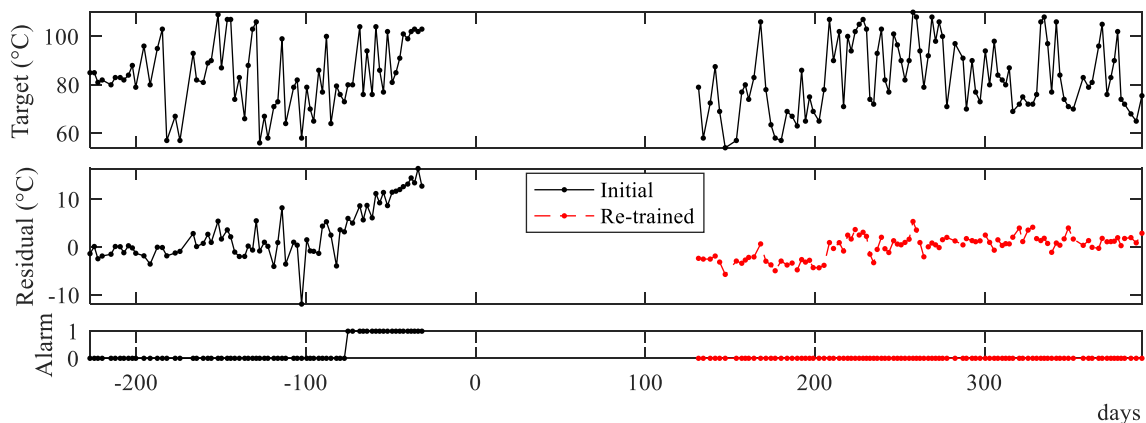


Figure 2. Detection of a gearbox problem with the time axis referring to the replacement date (temperature A, turbine 12).

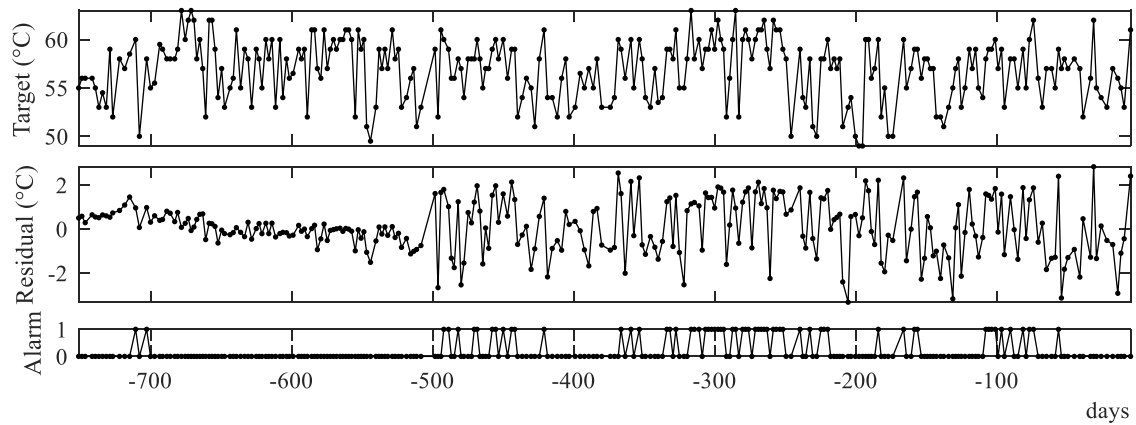


Figure 3. Detection of a gearbox problem with the time axis referring to the replacement date (temperature B, turbine 2).

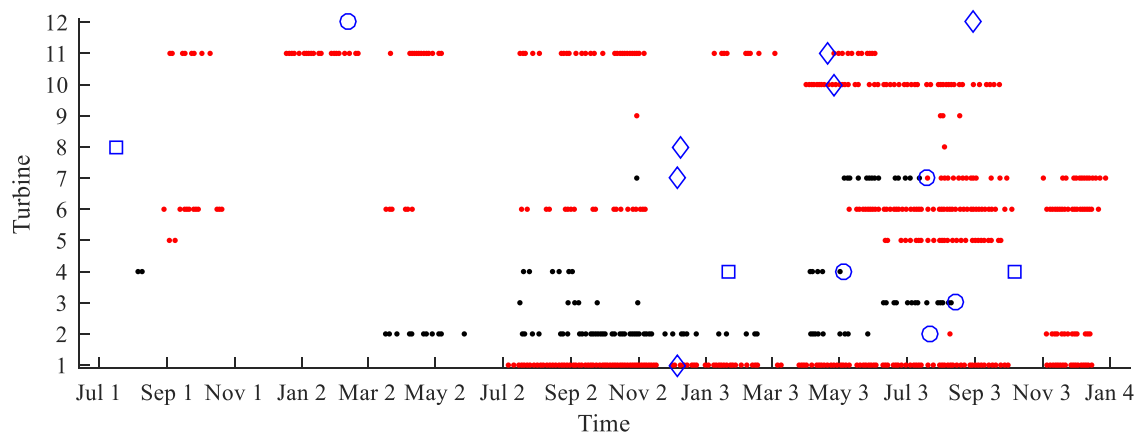


Figure 4. Alarms for gearbox problems in all turbines based on temperature B. Unrelated alarms are marked red, gearbox replacements with a circle, generator and bearing replacements with a square and asterisk, respectively.

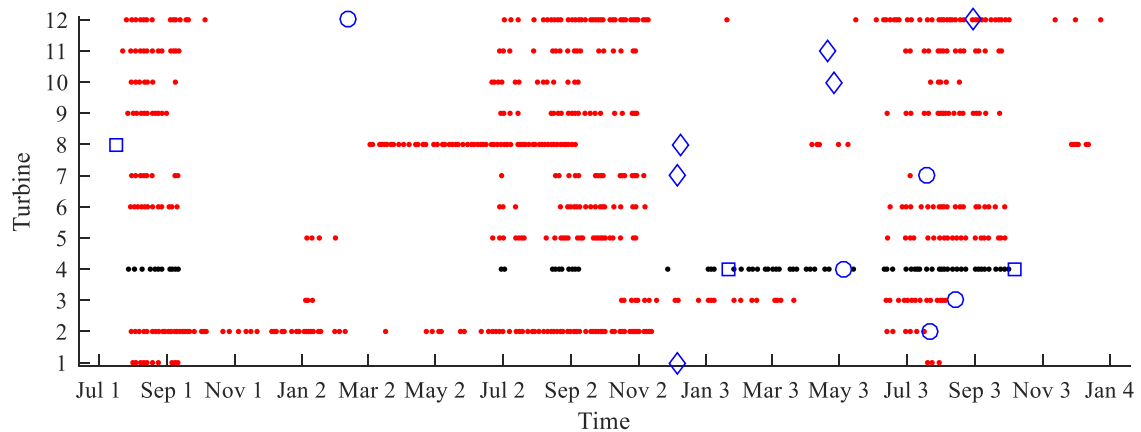


Figure 5. Alarms for generator problems in all turbines based on temperature C. Unrelated alarms are marked red, generator replacements with a square, gearbox and bearing replacements with a circle and asterisk, respectively.

4.2. Operational statistics

The analysis of the defined performance parameters showed that the whole farm is affected by changing operation during the whole 2.5 years of data as the parameter values from all turbines clearly vary with

time. As there is no common pattern, it is most likely that the reported replacements of gearboxes, generators and bearings have diverse causes and failure modes. Examples are given in figures 6 and 7 for selected dates with highlighted replacements happening in this month. The generator problem in turbine 4, figure 6a, seems to be related to relative high wind speed and accordingly rated power operation and high speed. Noticeably, the reactive power generation was exceptionally high in this time in several turbines including the failing one (average power factor of 0.9947). The bearing replacements, figure 6b and figure 7a, are found with various parameter values. Although most of the replacements show low or average parameter values, some are linked to high turbulence in operation. A high turbulence could also be the driver of the two gearbox replacements in figure 7b.

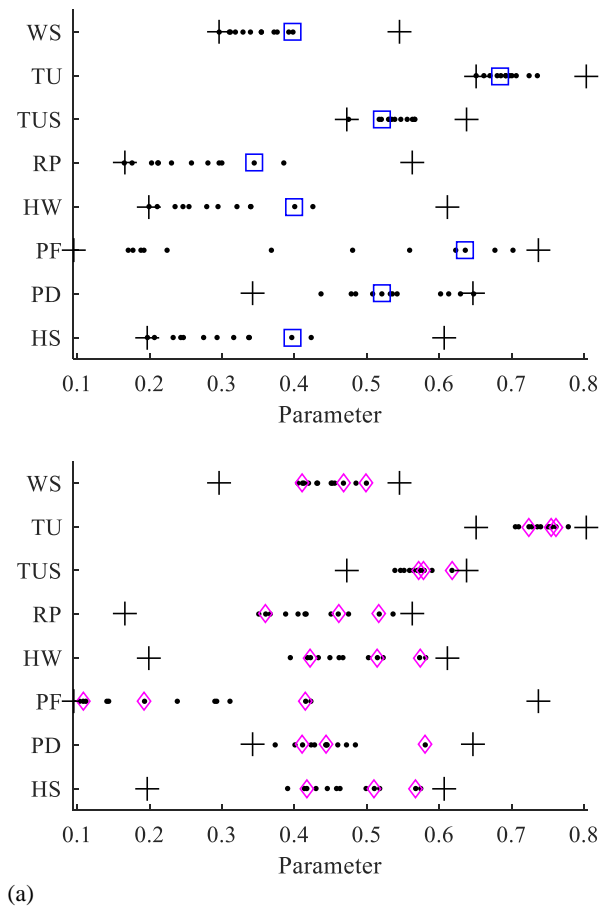


Figure 6. Performance parameters for all turbines in July year 1 (a) and December year 2 (b). Generator and bearing replacements marked with square and diamond, respectively. The extrema of the parameters from all months are marked with a plus symbol.

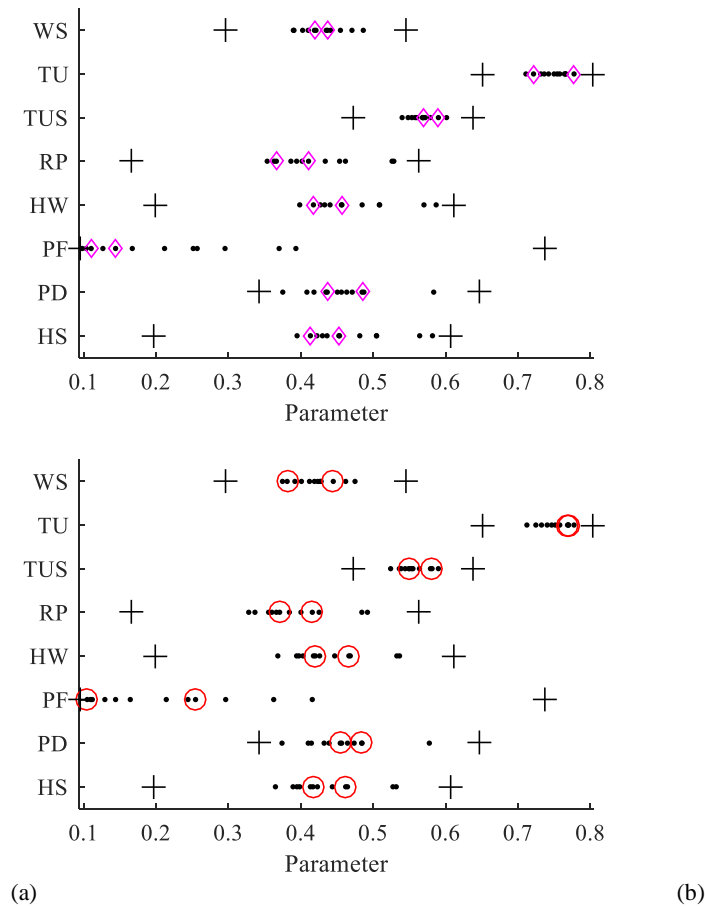


Figure 7. Performance parameters for all turbines in April year 3 (a) and July year 3 (b). Bearing and gearbox replacements marked with diamond and circle, respectively. The extrema of the parameters from all months are marked with a plus symbol.

5. CONCLUSION

Operational data from wind turbines could build an alternative and complement of dedicated vibration measurements. Model-based monitoring is a way to detect anomalies in the behaviour of wind turbine drive train temperature signals to detect mechanical degradation and possible failures. In contrast, the statistical analysis used in the Physics of Failure approach tries to identify turbines at risk by evaluating the damage drivers with performance parameters. A combination of the two approaches is proposed to increase the reliability of monitoring.

In a case study, both approaches are applied with the aim of finding early indications for several gearbox, generator and generator bearing replacements. In the model-based monitoring with ANNs and thresholds based on the residual distribution from training, early alarms for all gearbox replacements are issued. Similarly, generator problems in one turbine show up if using another temperature signal. However, many unrelated or possibly false alarms in turbines without reported problems of this type reveal challenges in getting reliable monitoring. The evaluation of the performance parameters results in the conclusion that different damage drivers and failure modes were involved. Particular high values in turbulence, reactive power generation and wind speed are found to correlate with some of the failed turbines. Although the properties of the case study limit the capabilities of both approaches, it can be seen that the combination of model-based monitoring and statistical analysis of SCADA data increases the knowledge of the system's condition.

In future works, the performance parameter values of this farm shall be compared to farms with similar settings. However, a thorough evaluation of the benefit of combining the two monitoring approaches will need better case data with a bigger farm size, more fault-free turbines and sufficient documentation.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642108 (project AWESOME, <http://awesome-h2020.eu/>).

REFERENCES

- [1] Carroll, J., McDonald, A., McMillan, D. (2015) Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy* 17:1107–19. doi: 10.1002/we.1887.
- [2] Reder, M., Gonzalez, E., Melero, J.J. (2016) Wind Turbine Failure Analysis - Targeting current problems in Failure Data Analysis. *J Phys Conf Ser* 753. doi: 10.1088/1742-6596/753/7/072027.
- [3] Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T.A., Ekemb, G. (2014) Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges. *Energies* 7:2595–630. doi: 10.3390/en7042595.
- [4] Qiao, W., Lu, D. (2015) A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis - Part II: Signals and Signal Processing Methods. *IEEE Trans Ind Electron* 62:6546–57. doi: 10.1109/TIE.2015.2422394.
- [5] Qiu, Y., Feng, Y., Tavner, P., Richardson, P., Erdos, G., Chen, B. (2012) Wind turbine SCADA alarm analysis for improving reliability. *Wind Energy* 15:951–66. doi: 10.1002/we.513.
- [6] Gonzalez, E., Reder, M., Melero, J.J. (2016) SCADA alarms processing for wind turbine component failure detection. *J Phys Conf Ser* 753. doi: 10.1088/1742-6596/753/7/072019.
- [7] Feng, Y., Qiu, Y., Crabtree, C.J., Long, H., Tavner, P.J. (2013) Monitoring wind turbine gearboxes. *Wind Energy* 16:728–40. doi: 10.1002/we.1521.
- [8] Schlechtingen, M., Ferreira Santos, I. (2011) Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mech Syst Signal Process* 25:1849–75. doi: 10.1016/j.ymssp.2010.12.007.
- [9] Bangalore, P., Tjernberg, L.B. (2015) An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. *IEEE Trans Smart Grid* 6:980–7. doi: 10.1109/TSG.2014.2386305.
- [10] Sun, P., Li, J., Wang, C., Lei, X. (2016) A generalized model for wind turbine anomaly identification based on SCADA data. *Appl Energy* 168:550–67. doi: 10.1016/j.apenergy.2016.01.133.
- [11] Schlechtingen, M., Santos, I.F., Achiche, S. (2013) Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description. *Appl Soft Comput* 13:447–60. doi: 10.1016/j.asoc.2013.09.016.
- [12] Wang, Y., Infield, D. (2012) Supervisory control and data acquisition data-based non-linear state estimation technique for wind turbine gearbox condition monitoring. *IET Renew Power Gener* 7:350–8. doi: 10.1049/iet-rpg.2012.0215.
- [13] Tautz-Weinert, J., Watson, S.J. (2016) Comparison of different modelling approaches of drive train temperature for the purposes of wind turbine failure detection. *J Phys Conf Ser* 753:72014. doi: 10.1088/1742-6596/753/7/072014.
- [14] Gray, C.S., Watson, S.J. (2010) Physics of Failure approach to wind turbine condition based maintenance. *Wind Energy* 13:395–405. doi: 10.1002/we.360.

On the Feasibility of Inferring the Applied Mechanical Loading of a Conveyor System Test Rig from Monitored System Parameters

Owen Freeman Gebler^{1,2}, Ben Hicks¹, Andrew Harrison¹, Matt Barker²

¹University of Bristol, University Walk, Bristol BS8 1TR, UK

²Stirling Dynamics, 26 Regent St., Bristol BS8 4HG, UK

ABSTRACT

Conveyor systems are ubiquitous throughout industry, providing materials transfer functionality across a wide range of applications. High availability requirements coupled with flexible modes of operation demand that maintenance of systems be conducted as efficiently as possible. In this vein, this paper presents an investigation into the feasibility of observing the presence of mechanical loading via the responses of a range of monitored system parameters, with a view to understanding the relationship between system operation and health.

A test rig is used to enable the emulation an industrial conveyor system's dynamics, and a wide range of sensors are employed to enable a comprehensive parameter set to be observed. A range of mechanical loading mechanisms are used to replicate the presence of loads typically experienced by a conveyor system, and a number of test scenarios are conducted, comprising both isolated and combined loading scenarios.

It was observed that, when applied in isolation, the presence of axial, radial and torsional loads applied to the rig can feasibly be identified from the combined response of a unique subset of system parameters. However, as more complex modes of loading are introduced, both in terms of profile and combinations of loads, less clarity in responses can be observed, with significant cross-coupling of effects present, suggesting that isolating conveyor loading within an industrial environment is likely to require leveraging of a wide range of parameters and state-of-the-art signal processing techniques.

Keywords: condition monitoring, maintenance, diagnostics, conveyor systems, sensors

Corresponding author: O. Freeman Gebler (email: o.freemangebler@bristol.ac.uk)

1. INTRODUCTION

Conveyor systems are fundamental industrial assets, which, at the highest level, serve to provide the function of material transfer. Their typically simple design and construction enables use throughout a wide range of industrial applications, with lengths ranging from a few metres to kilometres. This utility enables single systems to potentially be utilised for multiple, different purposes throughout its operational life, resulting in exposure to a wide range of loading conditions, potentially beyond those within their design envelope.

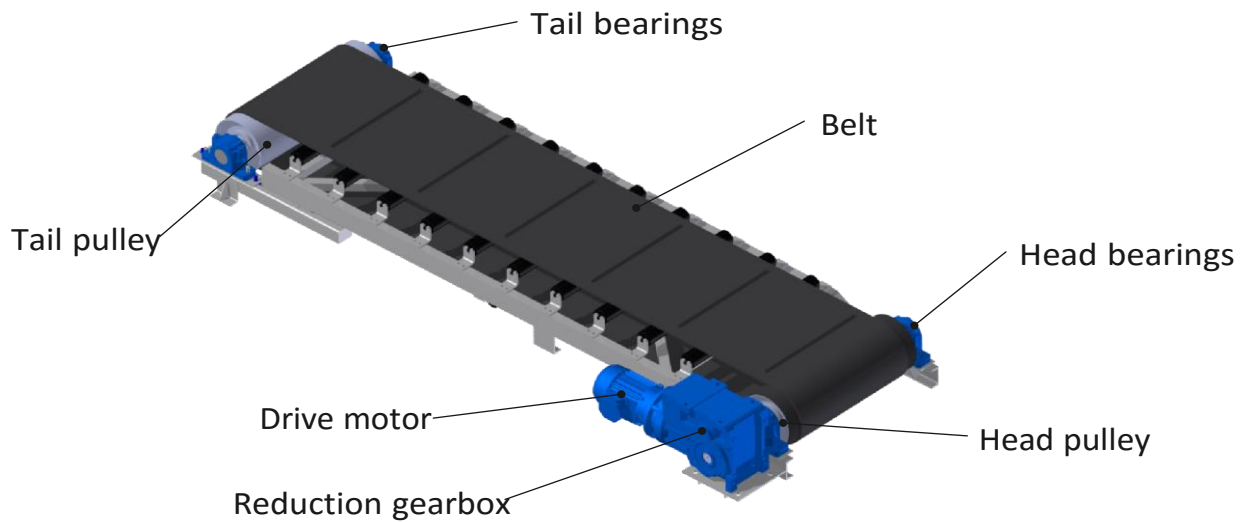


Figure 1: Layout of typical industrial conveyor system, with major components labelled.

Such variation in operational conditions will likely have implications for the form and frequency of maintenance activities required to support continued operation of a system. With systems typically subjected to very high utilisation demands and having severe penalties associated with lost operational time, it is key that maintenance activities can be planned and conducted efficiently, ensuring that only minimal downtime be incurred.

As such, there is a need to understand how the loading a conveyor system is subjected to during operation impacts upon the long-term health of the system. Accordingly, this paper presents the findings from a body of testing completed to characterise the response of system parameters to a range of loading conditions, using a bespoke test rig as a proxy for an industrial conveyor system.

A wider range of system parameters are selected for continuous monitoring throughout testing, to enable an evaluation of each's potential a proxy for indicating the presence and magnitude of applied loads, with a view to informing the specification of an industrial conveyor monitoring system.

2. METHODOLOGY

2.1. Conveyor Emulation Rig

To enable the characterisation of conveyor parameter responses to various loading scenarios a conveyor emulation rig (CER) was designed and constructed, allowing the dynamics of a typical industrial conveyor system to be replicated in a controlled environment. Within the CER a series of loading mechanisms are incorporated, enabling the application of axial, radial and torsional loading, either in isolation or simultaneously.

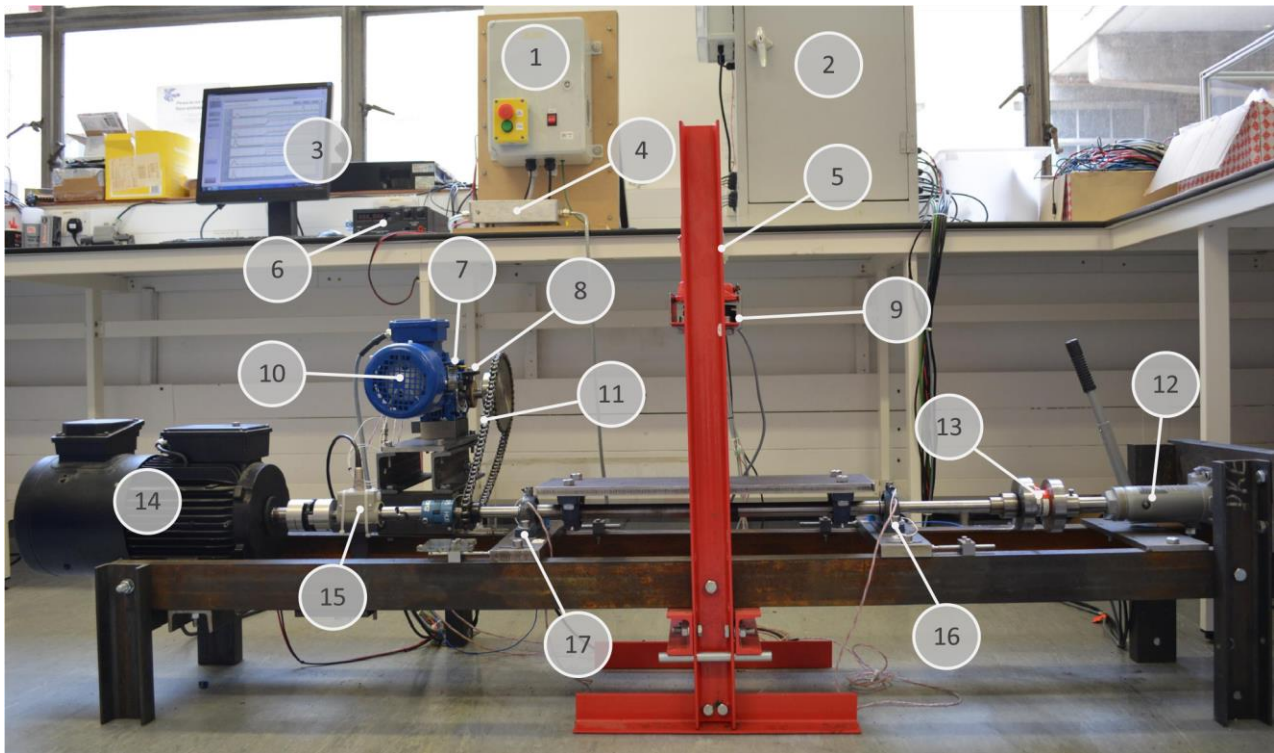


Figure 2: Overview of CER: 1 - VFD box; 2 - DAQ cabinet; 3 - Control PC and HMI; 4 - Power monitoring; 5 - Radial jack; 6 - DC injection power supply; 7 - Gearbox; 8 - Speed sensor; 9 - Radial loadcell; 10 - Drive motor; 11 - Drive chain; 12 - Axial jack; 13 - Axial loadcell; 14 - Braking motor; 15 - Torque transducer; 16 - Bearing 2; 17 - Bearing 1

A body of failure mode and effect analysis (FMEA) was conducted based upon an industrial conveyor system, enabling loads typically experienced by such a system to be identified, from which CER scenarios could be designed to enable the replication of these. It should be noted that the mechanisms of belt failure are not the focus of this research at present, therefore the CER omits a belt and instead uses the various loading mechanisms to emulate the presence of a belt.

Axial and radial loads are applied using dedicated hydraulic jacks, actuated manually, whilst torsional load is applied via a second braking motor, where the principles of direct current (DC) injection braking are employed. The strength of the torsional load generated is controlled by modulating the amplitude of the DC injected, thus by employing a programmable power supply more complex torsional load profiles can be generated e.g. saw tooth, quasi-sinusoidal.

Table 1: Summary of CER instrumentation

| | Monitored parameter | Sensor | Signal conditioning | Acquisition unit | Acquisition rate | Continuous/Periodic |
|---------------|-----------------------------|-------------------------|-------------------------------------|------------------|------------------|---------------------|
| Drive motor | Line voltages (VAB and VCA) | LEM LV25-P | Custom LP filter and amplifier unit | NI USB6211 | 10kHz /phase | C |
| | Phase currents (IA and IB) | LEM LA25-NP | Custom LP filter and amplifier unit | NI USB6211 | 10kHz /phase | C |
| | Casing temperature | PT100 | Industrial Interface E100 | | 2Hz | C |
| | Vibration | Dytran 3255A2 | NI 9234 | | 50kHz | P |
| Gearbox | Output shaft speed | Honeywell 103SR13A-1 | NA | NI USB6211 | Counter input | C |
| | Casing temperature | PT100 | Industrial Interface E100 | | 2Hz | C |
| | Audible noise | Bruel & Kjaer type 4117 | NI 9234 | | 50kHz | P |
| | Acoustic Emission | Mistras WD | Mistras 2/4/6C voltage amplifier | NI PCI9251 | 1MHz | P |
| Bearing 1 | Casing temperature | PT100 | Industrial Interface E100 | | 2Hz | C |
| | Vibration | Dytran 3255A2 | NI 9234 | | 50kHz | P |
| | Audible noise | Bruel & Kjaer type 4117 | NI 9234 | | 50kHz | P |
| Bearing 2 | Casing temperature | PT100 | Industrial Interface E100 | | 2Hz | C |
| | Vibration | Dytran 3255A2 | NI 9234 | | 50kHz | P |
| | Audible noise | Bruel & Kjaer type 4117 | NI 9234 | | 50kHz | P |
| Radial jack | Applied load | Tedea Huntleigh TH220 | Soemer LAC65.1 amplifier | NI USB6211 | 1Hz | C |
| Axial jack | Applied load | Novatech F210 | Vishay Nobel AST 3IS | NI USB6211 | 1Hz | C |
| Braking motor | Output power | Manson HCS-3302-USB | NA | | 2Hz | C |
| | Applied torsional load | HBM T5 | Vishay Nobel AST3IS | NI USB6211 | 1Hz | C |
| Environment | Ambient temperature | PT100 | Industrial Interface E-100 | | 2Hz | C |

To enable potential indicators of load to be explored a wide range of sensing is included within the CER (table 1). Monitored parameters were selected based upon a review of both extant literature as well as technical standards, enabling those parameters expected to present most sensitivity to load to be identified, whilst considering implementation factors such as cost and complexity also.

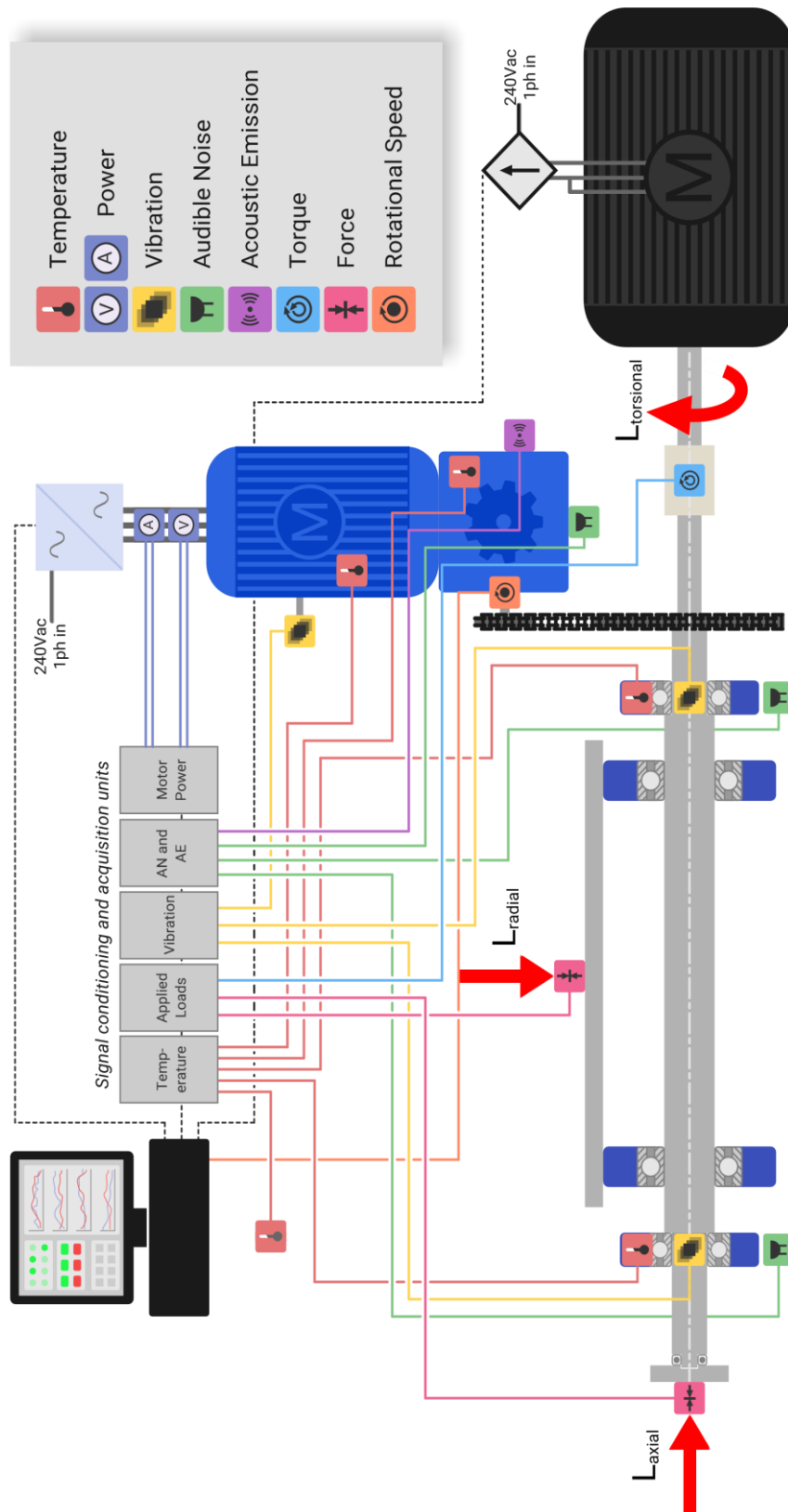


Figure 3: CER instrumentation schematic

2.2. Testing Regime

A total of 5 loading scenarios were investigated on the CER: axial only, radial only, torsional only, axial and radial, radial and braking. For axial and radial loads step profiles were applied only, while for torsional loads both step and square-wave profiles were applied, with the magnitudes in table 2 used for each.

Table 2: Summary of magnitudes of test variables used

| Variable | Abbrev. | Step 1 | Step 2 | Step 3 |
|----------------------|---------|--------|--------|--------|
| Axial Load (kN) | A | 2 | 4 | 6 |
| Radial Load (kN) | R | 1 | 2 | 3 |
| Torsional Load (Adc) | T | 2 | 5 | 8 |

After the application of each loading magnitude monitored parameters were allowed to reach steady state values prior to the application subsequent steps. In total, each scenario was conducted 3 times to enable the repeatability of observations to be investigated.

3. RESULTS

3.1. Isolated Loading Scenarios

When an axial load is applied to the CER in isolation the rotational speed of the gearbox output shaft reduces by approximately 1rpm per 2kN step of load. A corresponding increase in power consumption is not seen; with the VFD operating in open-loop V/f mode no attempt to compensate for the increased load on the drive is made.

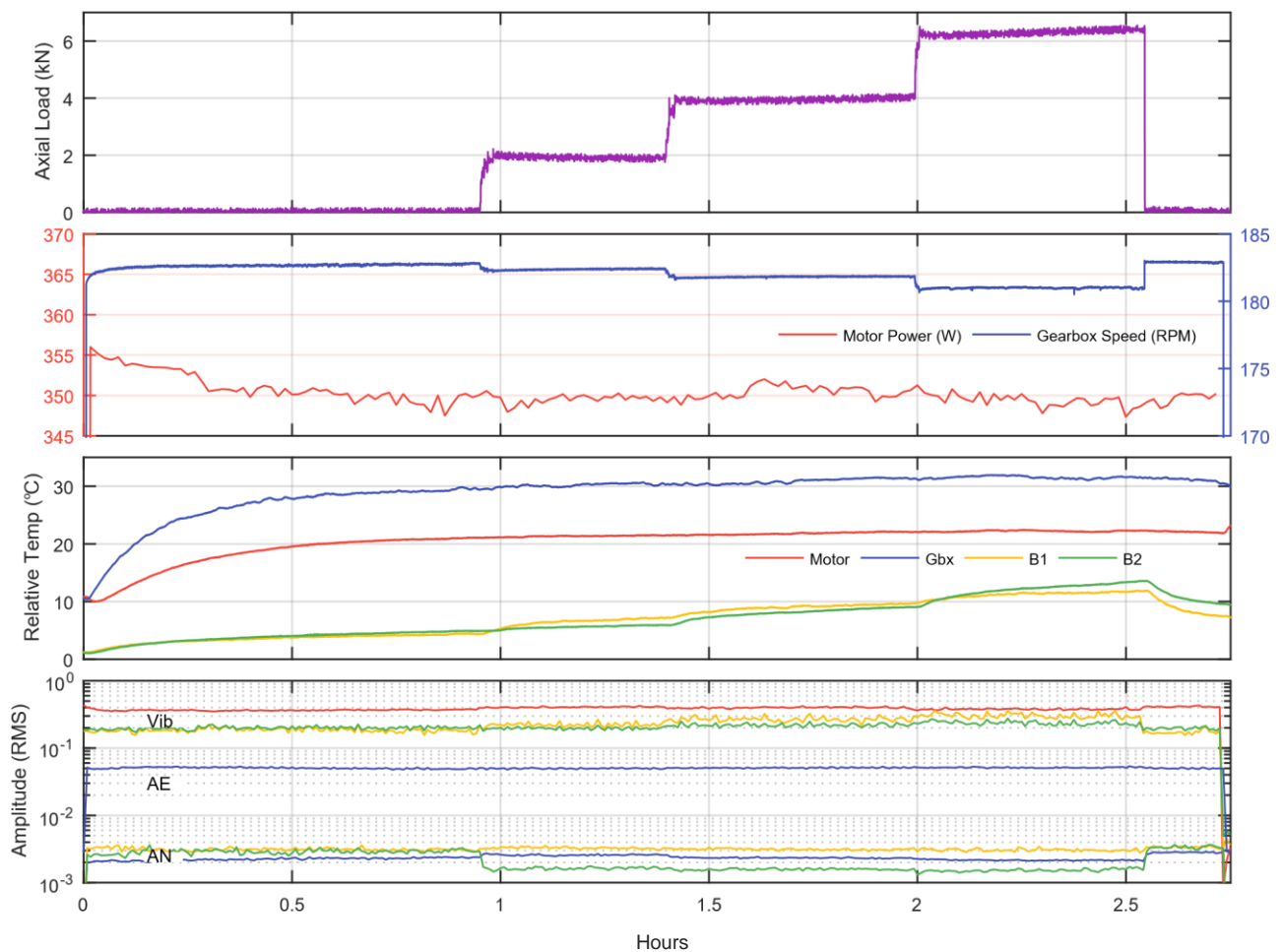


Figure 4: Axial loading characterisation test scenario

In general, little change is seen in motor and gearbox parameters in response to axial load. Vibration, audible noise and acoustic emission RMS values present minimal sensitivity to axial load, and the temperatures of each show only small increases from no load to maximum.

Drive bearings present more significant sensitivity to axial load; the vibration RMS values of both Bearing 1 and 2 show a clear increase corresponding to each load step applied, and the audible noise RMS value of Bearing 2 shows a sharp reduction as load is applied. The temperature of each bearing also increases in response to greater load. Initially, at steps 1 and 2 the absolute temperature of bearing 1 exceeds bearing 2 by 1-2°C, as a result of bearing 1 supporting the majority of the axial load. However, as step 3 is applied the temperature of bearing 2 can be seen to exceed bearing 1. This observation can be explained by the design of the CER; axial load is applied via a thrust bearing to minimise parasitic torsional load. Increasing the load subjected to the thrust bearing causes a greater level of heat to be generated, with thermal images suggesting a casing temperature of 70°C at 6kN (figure 5). This added heat is conducted down the drive and thus over time causes an ‘artificial’ increase in the temperature of bearing 2, which lies in closest proximity.

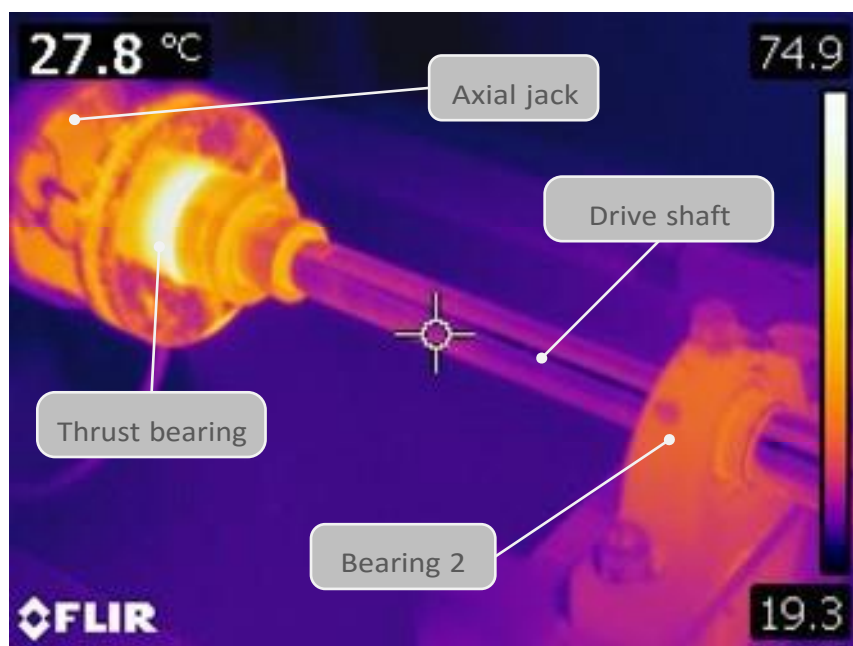


Figure 5: Thermal image of driveshaft and axial loading location during the application of 6 kN axial load-note maximum scale temperature of 74.9°C

Similarly to the application of axial load, when radial load is applied in isolation the CER motor and gearbox show minimal sensitivity; due to the chain drive configuration employed, axial and radial loads are not reacted significantly by the drive elements of the CER. No significant response to radial load can be observed in motor and gearbox speed, power consumption, temperature, RMS vibration or RMS acoustic emission across the applied range (figure 6).

In contrast, both drive bearings show significant sensitivity to radial load, with both present increased levels of AN RMS across the range. This is likely a consequence of a slackening effect on the chain, caused by deflection of the driveshaft in response to radial load. Bearing 1 can be seen to present a greater magnitude RMS value, likely as a result of its proximity to the drive chain. Both bearings also show an decrease in temperature under radial load, with Bearing 1 reducing by 1°C and Bearing 2 by 2-3°C, and a small increase in vibration RMS amplitude.

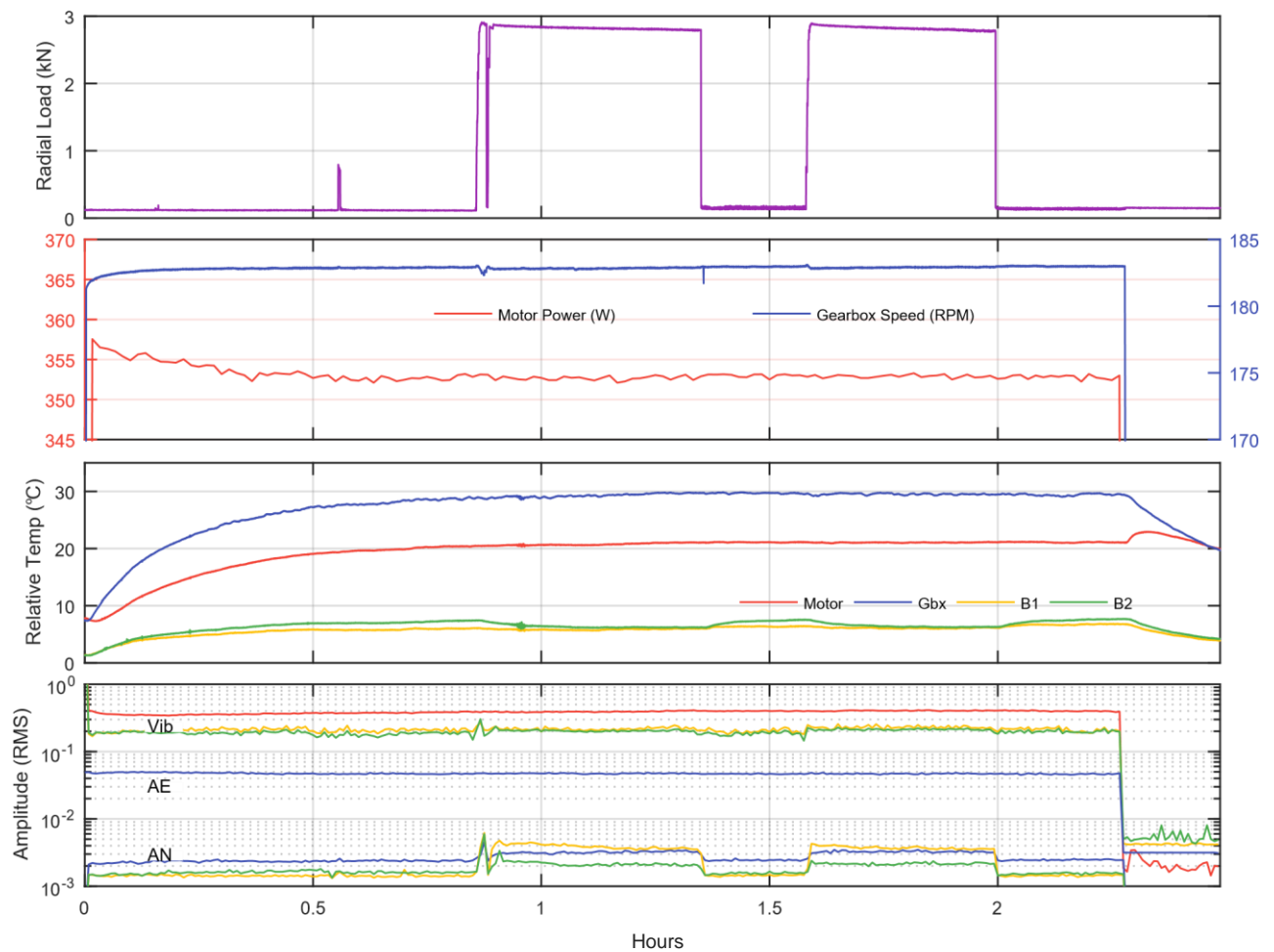


Figure 6: Radial loading characterisation test scenario

When torsional load is applied to the CER it is reacted primarily by the motor and gearbox, resulting in an observed simultaneous decrease in speed and increase in power consumption of 5rpm and 15W respectively at 2Nm load. Additionally, motor vibration RMS and gearbox AE and AN RMS values can also be seen to decrease as torsional load is applied, likely as a result of the reduction in operating speed.

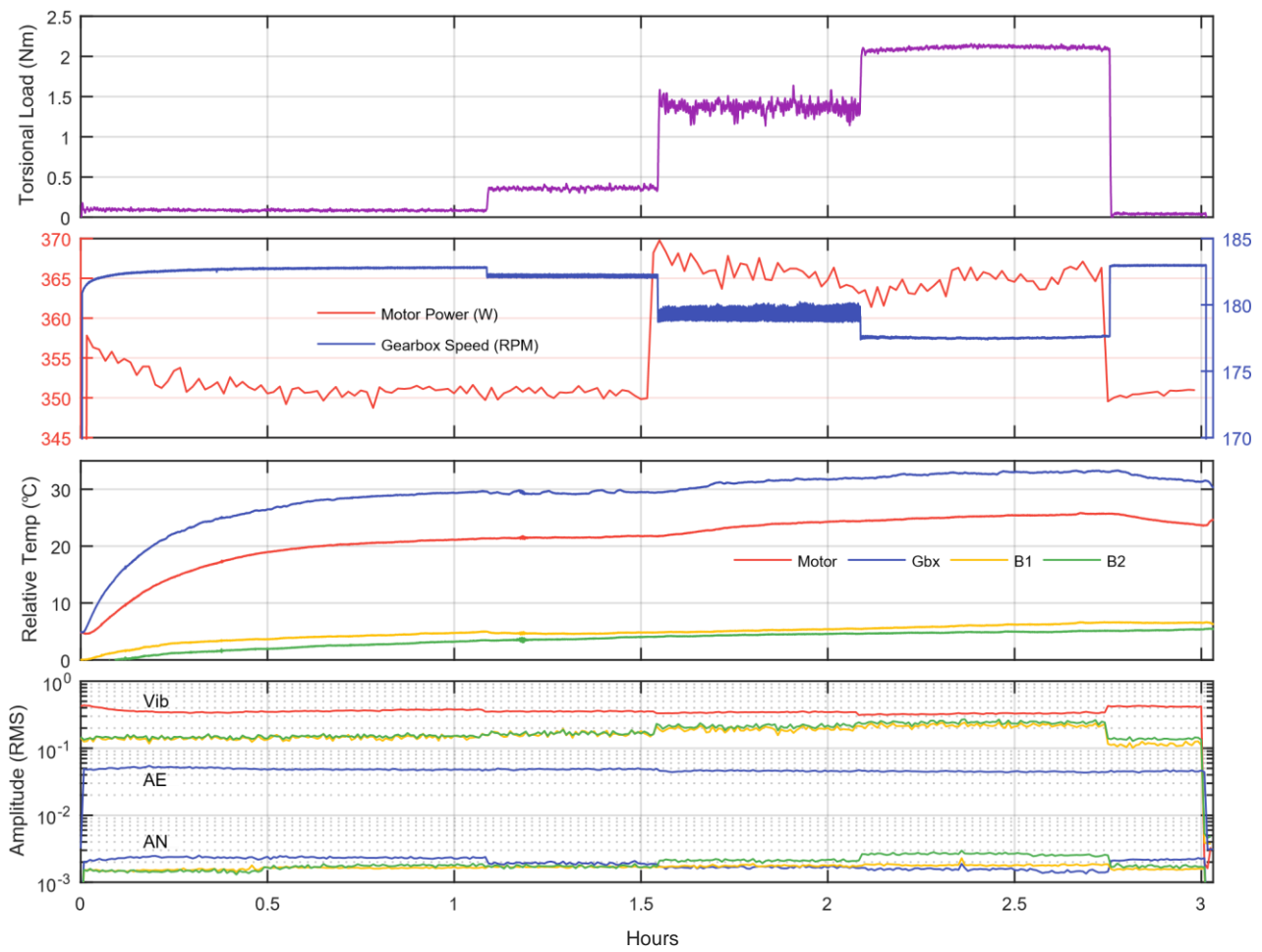


Figure 7: Torsional loading characterisation test scenario

No significant response to torsional load can be observed in the temperature of the drive bearings. Both do however show increased vibration and AN RMS values, likely caused by the slackening of the drive chain in response to the presence of counter-torque, resulting in Bearing 1 AN RMS being greater than that of Bearing 2 (figure 7).

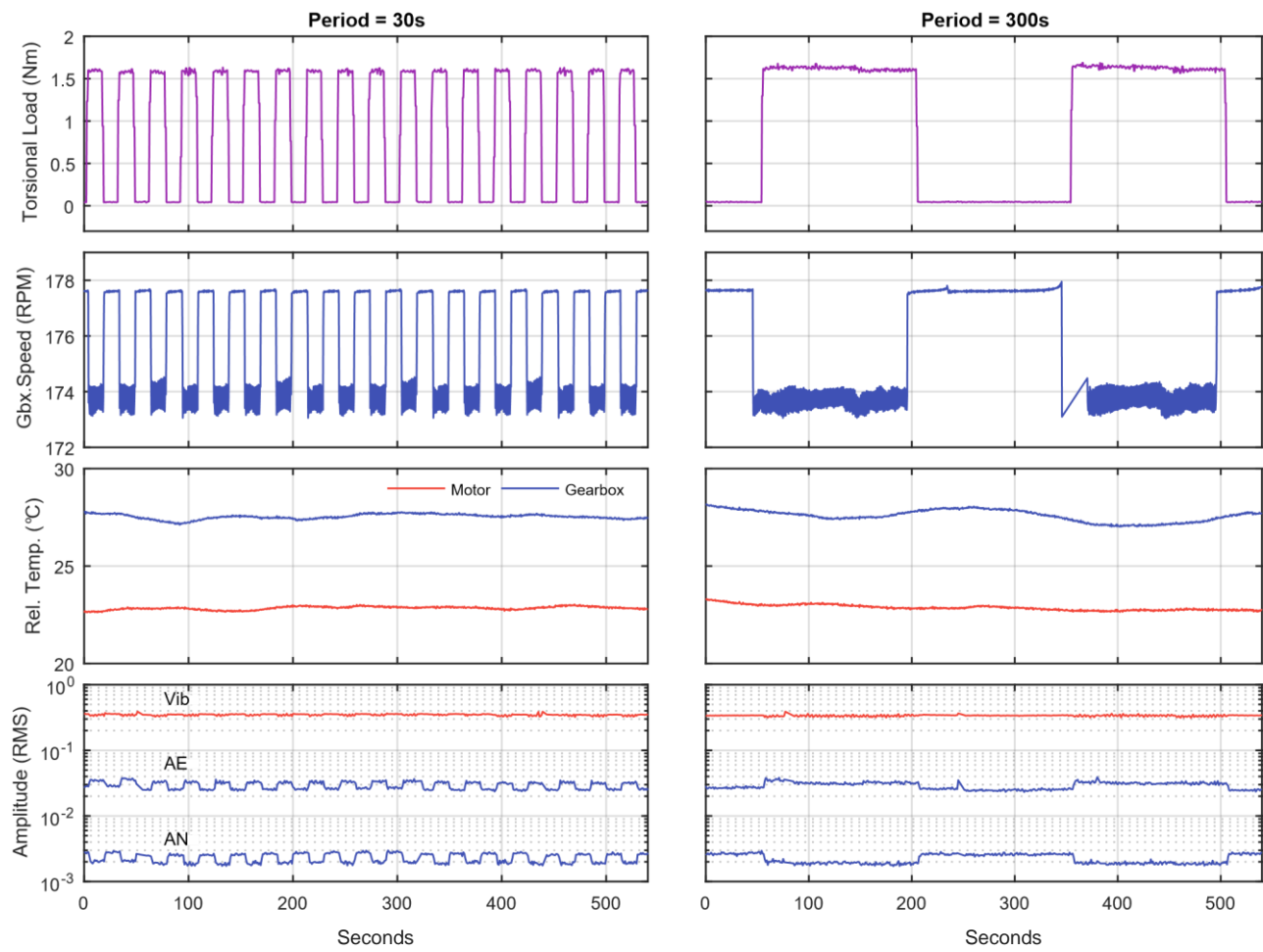


Figure 8: Torsional square-wave loading characterisation test scenario

3.2. Combined Loading Scenarios

When the CER is subjected to both axial and radial load simultaneously, as with isolated loading, motor and gearbox parameters present little sensitivity. A small reduction in power consumption and rotation speed is observed, along with minimal change in motor and gearbox vibration, AN and AE and temperature parameters.

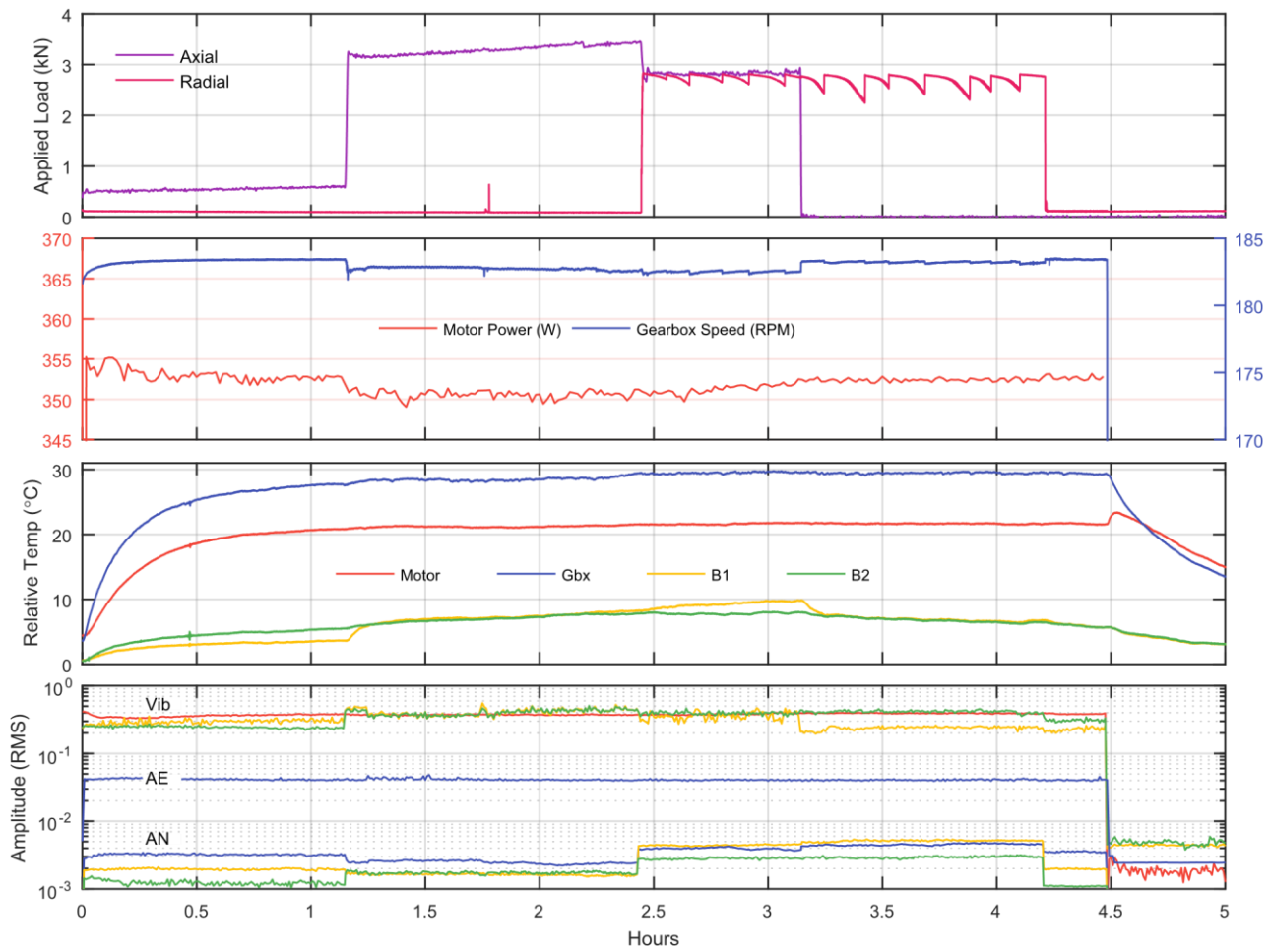


Figure 9: Combined axial and radial loading characterisation test scenario

As with isolated axial loading, initially after axial load has been applied the temperature of Bearing 1 increases whilst Bearing 2 remains relatively unchanged. However, when radial load is applied in addition to axial load the temperature of Bearing 2 begins to fall, as previously observed during isolated radial tests, whereas the temperature of Bearing 1 continues to increase until axial load is removed, after which Bearing 1 temperature falls also. Similarly, the vibration RMS values of the drive bearings present sensitivity to both modes of loading without an obvious trend. Initially, both bearings increase as the applied axial load is increased, however both fall as radial load is applied. As axial load is removed Bearing 1 falls to a level below that seen prior to the application of any load, whereas Bearing 2 shows little change, and when finally radial load is removed both values decrease further.

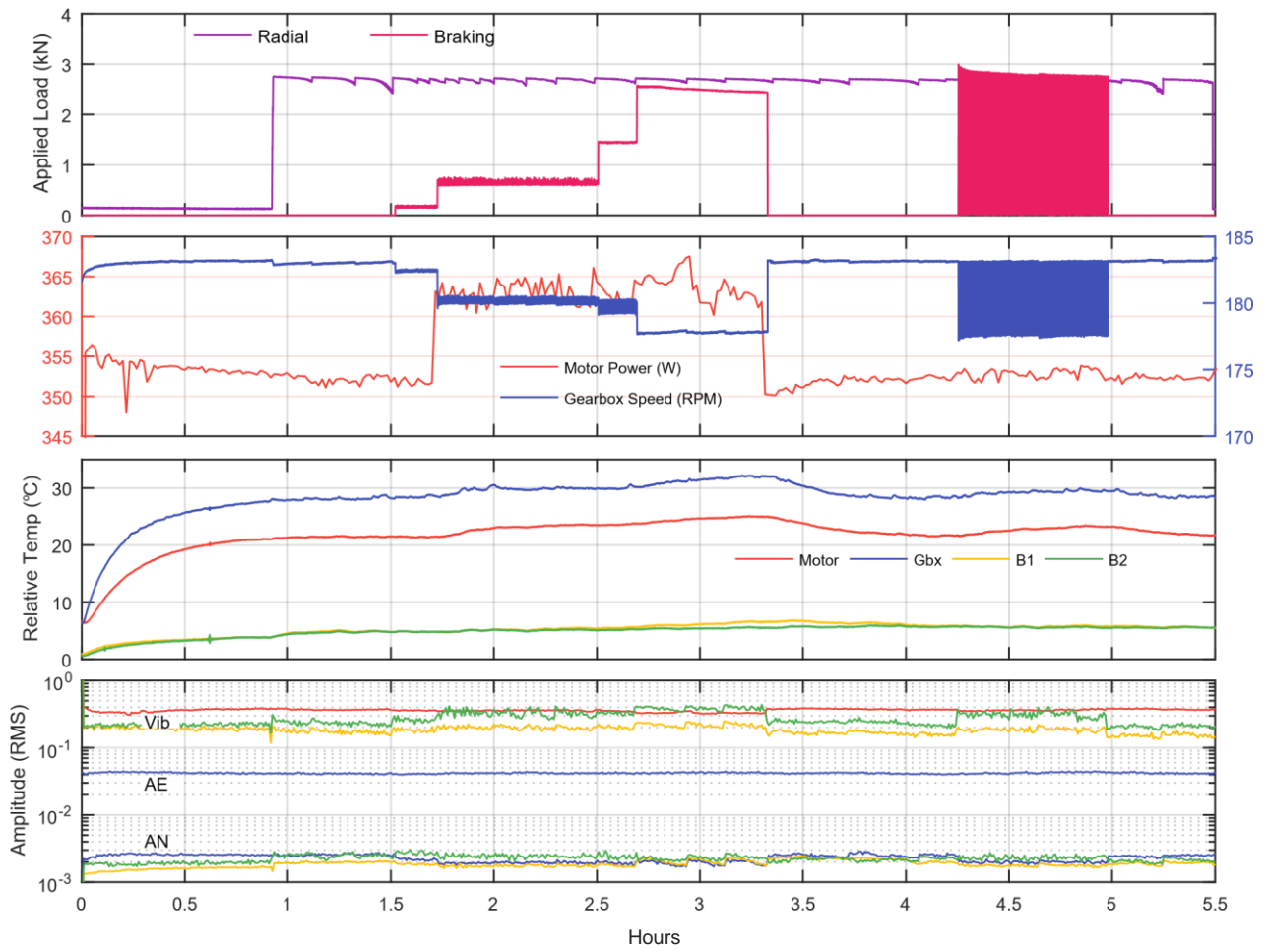


Figure 10: Combined radial and braking loading characterisation test scenario

A combination of radial and braking load applied to the CER results in a reduction in rotational speed of 5RPM at 2.5 kN of radial and 2.5Nm of torsional load, a similar reduction to that observed during isolated torsional load. Power consumption is increased also, however by a slightly lower magnitude than isolated torsional. The temperature of the motor and gearbox presents little sensitivity to radial loading, but when subjected to maximum torsional load increases by 4° C, again similar to isolated torsional loading. However, drive bearing temperatures do not present a reduction in response to radial load as previously observed, with a small increase in both seen when torsional load is applied. Motor vibration RMS and gearbox AE RMS present minimal sensitivity to combined radial and torsional loading, whilst gearbox AN RMS does show sensitivity, however no clear trend can be observed. The vibration RMS value of both bearings does present a significant increase in response to both modes of loading, with the magnitude of both exceeding levels observed during isolated loading. Similarly to gearbox AN RMS, bearing AN RMS values show sensitivity to both modes of loading applied simultaneously, however no clear trend can be observed.

4. DISCUSSION

Data acquired during operation of the CER suggests that, when applied in isolation, the presence and form of loading can feasibly be inferred via the combined response of specific system parameters (table 3).

Table 3: Summary of sensitivity of parameters to applied loading profiles

| Applied Loading | Sensitive parameters | Response |
|-----------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| Axial | B1 temp, B2 temp, B1 vib, B2 vib, B2 AN, | Bearing temperature and vibration RMS differential present - B1 greater than B2. Reduced B2AN RMS under load |
| Radial | B1 temp, B2 temp, B1 AN, B2 AN | Decrease in absolute temp and increase in AN RMS of both B1 and B2 |
| Torsional | Motor temp, gbx temp, motor vib RMS, gbx AN and AE RMS, motor power, gbx speed | Increased power drawn with decreased gbx speed. Increases in all other parameters |

Whilst in certain scenarios individual parameters can be able to indicate the presence of abnormal loading, by leveraging information from multiple parameters a greater degree of confidence in inferences can potentially be realised. For example, a change in operational condition due to an observed change in rotational speed can be detected directly from a single parameter, however, the root cause of this change may be radial or torsional loading. Only by considering the response of additional parameters such as power consumption and component temperatures can a more detailed inference be made, suggesting that more comprehensive sensing capability could support greater inferencing and confidence therein.

However, whilst these combinations of responses have presented potential as indicators of loading, the strength of the response of each varies significantly. Some have presented clear, quantifiable changes in response to the application of a loading condition, whereas others can only be considered as qualitative indicators of a change in operational conditions, with a lack of obvious, repeatable trends. As such, the level of inference potentially possible varies across the range of conditions.

Additionally, when multiple loading conditions are applied simultaneously, the ability to identify clearly each mode from the parameters available is reduced. Whilst changes in operation can be observed as loading conditions are changed, effects appear typically to be coupled, and as a result responses are not necessarily a summation of the responses observed during isolated loading.

As it's currently designed the CER can be seen to decouple the effects of certain loading conditions due to the physical layout of components. By driving the main shaft via a chain axial and radial loads applied are not significantly reacted by the motor and gearbox. Typically, an industrial conveyor system would be configured with a direct connection between the gearbox output and the head pulley, therefore it can be expected that in an industrial setting the response of components to loading may be even more tightly coupled than observed on the CER.

To attempt to decouple modes of loading, and thus enable isolation of individual modes of loading, additional system information accessed via signal processing techniques may provide value. For example, currently only time-domain analysis of parameters has been investigated, a large body of frequency-domain techniques are present within literature which may be able to provide valuable insight (see [1] for an overview). Alternatively, system parameter observability could be increased, supporting more detailed inferences, however, the impact of increasing DAQ hardware in an industrial setting must be considered. Additional hardware possesses an associated cost, both financial as well as in terms of system complexity and thus potentially reliability [2, 3], therefore the specific value of the information provided by each monitored parameter in supporting inferencing must be well understood.

However, ultimately techniques could, in the context of industrial conveyor systems, be used to support real-time monitoring of conveyor loading, with a view to enabling intelligent scheduling of PPM intervals, based upon the cumulative loading subjected to a system, as opposed to time in operational. Actions could be triggered upon the exceeding of thresholds, the levels of which could be determined empirically initially,

however, to mitigate the potential issues associated with defining appropriate thresholds [4, 5, 6] a more sophisticated and automated process may ultimately be required.

5. CONCLUSIONS

Data acquired during operation of the CER suggests that modes in isolation can feasibly be identified using specific combinations of parameter responses. Whilst individual parameter responses do, in some cases, present sensitivity to specific loading modes, greater confidence in inferences can potentially be achieved through the utilisation of information from multiple parameter responses in conjunction.

Additionally, as the complexity of loading was increased identifying each mode present proved increasingly challenging, with cross-coupling effects impacting the clarity of responses. To address this challenge it was suggested that increasing the range of parameters monitored, facilitated by either additional data acquisition hardware or advanced signal processing techniques, could support greater observability of responses to loading.

Accordingly, it can be concluded that further testing is required in order to improve understanding of the operational characteristics of the CER, as well as associated indicators, ultimately leading to the development of an approach to real-time monitoring of conveyor system loading ready for trialling within an industrial environment.

ACKNOWLEDGEMENTS

This work was supported by the EPSRC funded Industrial Doctorate Centre in Systems, University of Bristol and University of Bath (Grant EP/G037353/1) and the company sponsor Stirling Dynamics.

REFERENCES

- [1] A. K. S. Jardine, D. Lin, and D. Banjevic. "A review on machinery diagnostics and prognostics implementing conditionbased maintenance". In: *Mechanical Systems and Signal Processing* 20.7 (Oct. 2006), pp. 1483–1510.
- [2] O. Freeman Gebler et al. "Towards the implementation of a predictive maintenance strategy : Lessons learned from a case study within a waste processing plant". In: *European Conference of the Prognostics and Health Management Society 2016*. 2016, pp. 1–17.
- [3] W. W. Tiddens, A. J. J. Braaksma, and T. Tinga. "The Adoption of Prognostic Technologies in Maintenance Decision Making: A Multiple Case Study". In: *Procedia CIRP* 38 (2015), pp. 171–176.
- [4] E. Bechhoefer and A. P. F. Bernhard. "A generalized process for optimal threshold setting in HUMS". In: *IEEE Aerospace Conference Proceedings Ci* (2007).
- [5] M. Sawicki et al. "An automatic procedure for multidimensional temperature signal analysis of a SCADA system with application to belt conveyor components". In: *Procedia Earth and Planetary Science* 15 (2015), pp. 781–790.
- [6] P. K. Stefaniak et al. "Diagnostic Features Modeling for Decision Boundaries Calculation for Maintenance of Gearboxes Used in Belt Conveyor System". In: *Advances in Condition Monitoring of Machinery in Non-Stationary Operations. Applied Condition Monitoring*. Ed. by F. Chaari et al. Vol. 4. Springer International Publishing, 2016, pp. 251–262.

Assessing Uneven Milling Cutting Tool Wear using Component Measurement

Z. Ahmed ^{1,2}, P. Prickett ¹, R. Grosvenor¹

¹ Cardiff School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

² Department of Mechanical Engineering, AL- Anbar University, Iraq

ABSTRACT

Tool wear is a complex phenomenon inherent in any cutting process. Cutting tool wear monitoring is therefore deployed in CNC milling to support machining operations in order to plan tool changes and avoid economic losses. The application of tool life management strategies can lead to premature removal of healthy tool or the continued use of a dangerously worn tool. This has led to the investigation of more appropriate strategies. Depending upon the nature of the sensor technology deployed tool wear monitoring methods are categorized as being either direct or indirect. The benefits and challenges to machine tool users of both approaches are subject to a body of ongoing research.

In this study, a series of milling machining tests were performed in order to allow the confirmation of the presence of uneven tool flank wear. This was enabled by the indirect assessment of the tool condition by utilising a Coordinate Measurement Machine (CMM) to accurately measure the workpieces. Using a defined machining process with set cutting parameters each workpiece was machined to produce eight off 40 mm cylindrical holes; in this manner using four workpieces a series of 32 holes were machined. Each cylinder was machined using four separate cuts, at increasing depths, producing four identifiable sections.

Each section was measured and the form of the geometry produced was established. After assessing the diameters of all the sections for each cylinder, the presence of uneven flank wear was confirmed and the levels obtained. This is related directly to the differing amount of metal removed by the cutter during the established cutting cycle. The same processes was undertaken using three different sets of cutting parameters. The analysis showed the CMM to be a reliable basis for the measurement of uneven tool wear based upon the geometry of the component.

Keywords: Tool Wear, Coordinate Measure Machine CMM, Component Feature Geometry, spindle motor load monitoring.

Corresponding author: Z. Ahmed (AhmedZJ@cardiff.ac.uk and Zinah_jummah@hotmail.com).

1. INTRODUCTION

In this investigation the indirect assessment of the uneven tool flank wear was undertaken based upon monitoring the changes in the cutting process as indicated by component features. CMM based measurement of component feature geometry is shown to be capable of measuring tool wear indirectly. The measurement of the changes in the geometric features and form and the surface metrology of the machined component can be used in preference to assessing the state of the cutting edge of the cutting tool directly. This is shown to be particularly relevant in the context of uneven tool wear. The intention of this study is to inform ongoing tool wear monitoring research for which an understanding of the level and nature of the tool wear arising is vital. The intention is to then build a more accurate tool wear monitoring approach based upon this knowledge.

Monitoring the condition of cutting tools and tool life prediction plays an important role in improving machine productivity. It can enable and support techniques aimed at maintaining the quality and integrity of the machined part, minimizing material waste and reducing the cost. To be most effective it is necessary to

be able to observe tool wear during the material removal process. The wear of the cutting tool, which develops due to the dynamic interactions between the cutting tool and the workpiece, results in a reduction in the quality of the machined parts and the associated reduction in productivity. This may be related to reduced cutting speeds and less than optimal machining processes, which in turn may mean more parts may be rejected and/or have to be possibly reworked. Of the different types of tool wear the most commonly occurring in end milling are cutter flank wear (VB) and crater wear. Of the two flank wear is widely used as tool life criterion, with the established tool wear versus cutting time curve following the general form shown in figure 1. This is because flank wear in particular can have adverse effects on the final surface finish quality and the dimensional accuracy of the component. It is also perhaps the most convenient tool related property to measure, as it can be discerned by inspection of the machine artifact.

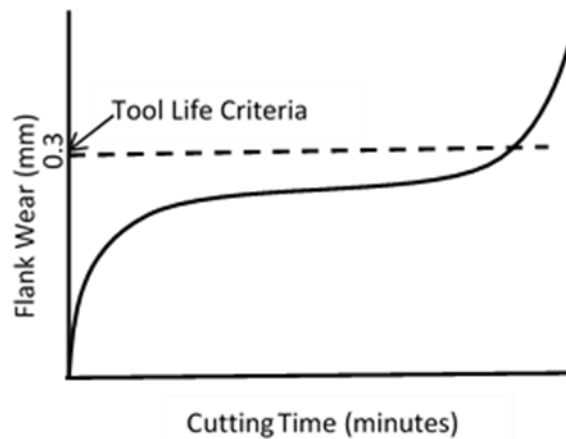


Figure 1. The relation between flank wear and tool life criteria

It is normal practice to assert that a tool should be considered to have reached the end of its useful life when flank wear has been attained to result in a specified dimension. According to ISO 8688-2 recommendations for both roughing and finishing cuts, for end milling cutters the end of useful tool life can be assumed to occur when tool wear levels equal to 0.3 mm for uniform wear and equal to 0.5 mm for a non-uniform (localized) wear arise [1]. These values are used in this work as an indication of the presence of unacceptable levels of tool wear.

The earliest effective approach to determine tool life for a given cutting speed was proposed by Taylor 1906 [2]. This approach suggested that, for progressive wear, the relationship between the time to tool failure for a given wear criterion and cutting speed was of the form:

$$V.T^n = C \quad (1)$$

Where: V is the cutting speed (m/min) and T is the tool life. This is normally measured in the most relevant time base (minutes). In this simplest of forms the constants n and C are defined for the particular combination of tool and workpiece material combinations and other machining variables such as feed rate and depth of cut. Values for n and C can be obtained from standard tables for different workpiece materials and different cutting tools. However, this equation does not include the effect cutting tool geometry, cutting feed, depth of cut and is limited to a certain range of speed [3]. Taylor's extended equation addresses this:

$$V^{\frac{1}{n_1}} f^{\frac{1}{n_2}} d^{\frac{1}{n_3}} T = C \quad (2)$$

Where the exponent n_1 , n_2 , and n_3 are constant and determined empirically. f is the feed (mm/rev.) and d is the depth of cut (mm). Since the provision of this, and similar considerations, various methods to calculate tool wear and tool life have been proposed. However, none of these equations can be applied with absolute certainty due to complex nature of the machining process [4]. The application of tool life management

strategies based upon these equations is difficult as machining conditions can change and the nature of the metal removal process depends upon the geometry of the component and the selected tool path. Part of this consideration is then related to uneven tool wear which reflects the nature of the cutting undertaken by a tool during its life. Thus it is likely that in real-life conditions the unguarded application of tool life calculation can lead to premature removal of healthy tool or the continued use of a dangerously worn tool.

Tool wear measurement can be broadly split in to two types, direct and indirect methods [5]. Direct measurements can be fast and accurate. They entail measuring the actual wear, using different methods including; optical measurement, radio-activation analysis and electrical resistance measurements. However, the direct measurement of tool wear during machining operations is difficult. Alternatively, indirect measurements may be online (or in process) and use machining process signals, such as cutting force, acoustic emission, sound, vibration [6] and current power for various drives [7]. There are other indirect measurements that are basically offline and relate to workpiece condition, including the measurement of the changes in machined component dimension or geometric form, the value of the volume of metal removed and component surface finish and/or roughness. However, in practice, tool wear is a very much more complex problem and it is potentially simple to determine tool wear quantitatively but is complicated in a practical context [8].

A considerable amount of research has been directed towards indirect measurements the tool wear based upon measuring the changes in geometric form or machined component dimension such as cylinder form and quality [9-15]. Little or no research into the quality of milling cylindrical features, such as the holes used in this work, has been previously published. The intention of this research is to explore how it is possible to under or overestimate the levels of tool wear when adopting a “traditional” method to calculate the average cylinder diameter as the measure of the actual tool wear. By simply averaging features that are measured the offset tool wear based on the component measurement will be inaccurate. Most importantly it is possible that the critical amount of tool wear especially for the bottom section of the cutter will be underestimated. This can mean that the machine tool operator will carry on machining using a cutter that is near to or exceeding the end of its life. It is also likely that the geometric form of a component machined using this cutter will be less than optimal.

2. EXPERIMENTAL UNEVEN TOOL WEAR ASSESSMENT

The aim of the experiments performed within this initial work was to establish and verify the approach adopted to form and feature measurement, which was designed to directly measure tool flank wear. In this work, the tool wear estimation method was developed based on component metrology. This approach was established to enable the consideration of the effect of machining conditions on tool wear. The key concept the method is to employ the appraisal of the form and dimension of features of a milled cylindrical cavity to quantify and classify the tool wear status.

2.1. Milling Machine

The initial experimental work was performed on a Mazak Vertical Centre Smart 430A (MVCS). The MVCS's ability to machine in three-axes enables the production of complex components and shapes. Using the full range of available spindle speeds, up to the maximum 12000 rpm, allows for a broad range of cutting parameters. Another feature that made the MVCS a suitable machine was its capacity to hold multiple tools and to undertake workpiece setting using on-machine probing. As the complete testing of tools needed more than one workpiece, this function allowed the cutting tool to remain in the machine without having to be removed and reset. This process could have potentially caused the tool setting to be different, thus affecting results.

2.2. Workpiece and Cutting Tool Material

In this study, a bright mild steel workpiece was used as the machining material. It was milled by using a high-speed steel 16 mm diameter 4-flute end mill cutter. The machining of bright mild steel is challenging due to the mechanical properties of hardness, tensile and yield strengths. It is also desirable to machine at higher cutting speeds and feeds while maintaining a good surface finish. The particular combination was

selected to induce tool wear on a realistic but accelerated basis. Figure.2 represents the workpiece condition after machining. The dimensions of the test piece were 125 mm x 220 mm x 25 mm. There are eight 40 mm diameter holes, numbered 1 to 8. The test piece design was intended to highlight the effect of tool wear via the measurement of the dimension of the machined features using a CMM.

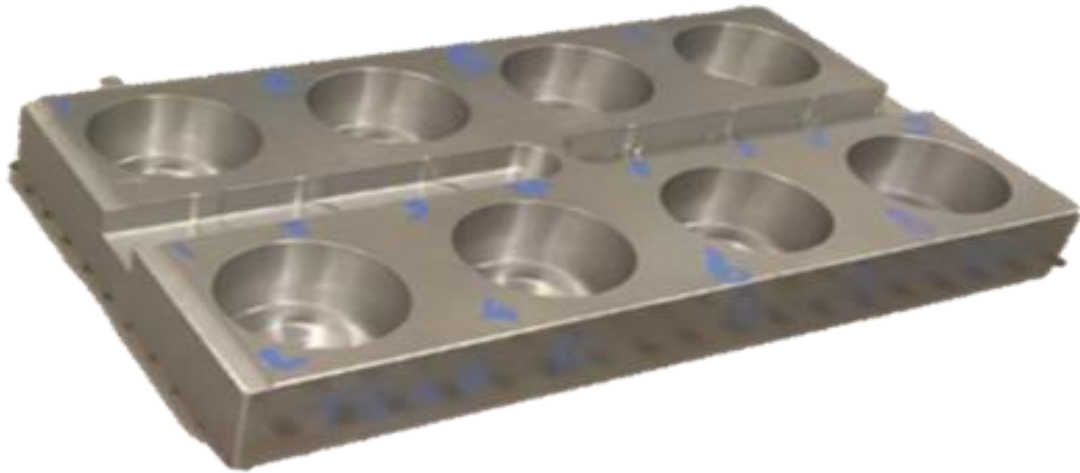


Figure 2. Eight-cylinder test piece

The conditions for the initial tests were selected by taking the recommended cutting speed for a milling operation. The cutting conditions chosen for this initial study were: spindle speed 1035 rpm, feed rate $258.7 \text{ mm} \cdot \text{min}^{-1}$ and cutting speed $52 \text{ m} \cdot \text{min}^{-1}$.

2.3. Experimental Procedure

At the start of each test cycle, a new cutter was used to machine the first 5mm deep cylinder C1 as shown in figure 3. A sequence of three further cylinders were then machined to depths of 10 mm cylinder C2, 15 mm cylinder C3 and 20 mm cylinder C4, again shown in figure 3. This sequence was repeated in each of the eight locations on a workpiece, as indicated in figure 2. On the basis that the material removed in machining Hole 1 Cylinder 1 was using the new tool, this hole was used as the reference hole. It is very important criterion because the diameter of the other holes at different depths and across different workpieces could be compared with it and the measured differences could then be treated as the tool flank wear.

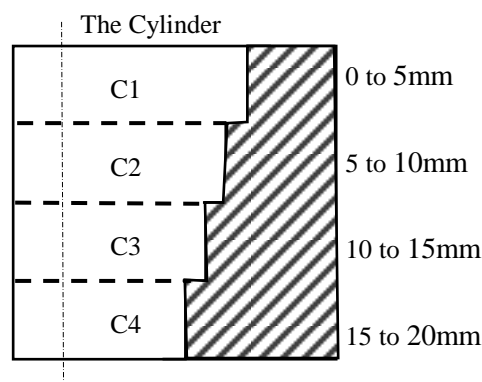


Figure 3. Cylinder measurement depth dimensions

The method adopted was aimed at indirectly measuring flank wear of the major cutting edge of the tool based upon feature metrology. In this set-up each cylinder was machined following the sequence shown in figure 4. The cycle was designed to remove a 5mm deep section of the 40 mm cylinder following a sequence of plunge-linear move-circular cut cycle, as indicated. Each cutter was used to produce eight times 40 mm holes that were machined in four workpieces. Each cylinder was then measured at each of the

four depths and an average diameter was established for each separate 5mm section cylinder. Using the measurements obtained at the four different depths an average diameter was established. The values of the average diameter for each of these four cylinders was then calculated, based on the coordinate measurements obtained. After assessing the cylinder diameters of all the holes and calculating flank wear, it was possible to produce plots from which underlying trends could be determined.

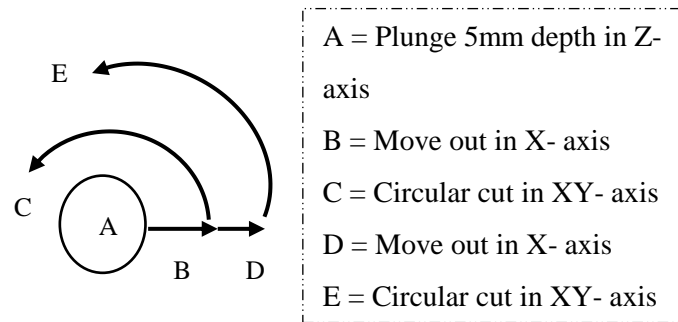


Figure 4. The Diagram of Cutting Operation for Each Cylinder

The CMM deployed in this research operated using a Renishaw 5-axis controller with state of the art measurement, sensor and control technology. The use of this 5-axis system enabled the accurate measurement of cylindrical features using a circular scan of the inside of the designated cylinders at the controlled depths indicated in figure 3. A cylinder feature was then constructed for each pair of measurements to provide the cylinder data used to produce figure 5. The application of this measurement process provided a precision of $<0.1 \mu\text{m}$. The deployed Revo system is also capable of surface finish measurement, which will be utilised in the later stages of this research.

3. RESULTS AND DISCUSSION

Figure 5 presents the initial results from the CMM measurements of the four diameters, D1, D2, D3 and D4 corresponding to the four cylinders C1, C2, C3 and C4. The calculated values for the average whole diameter D ave. is also presented.

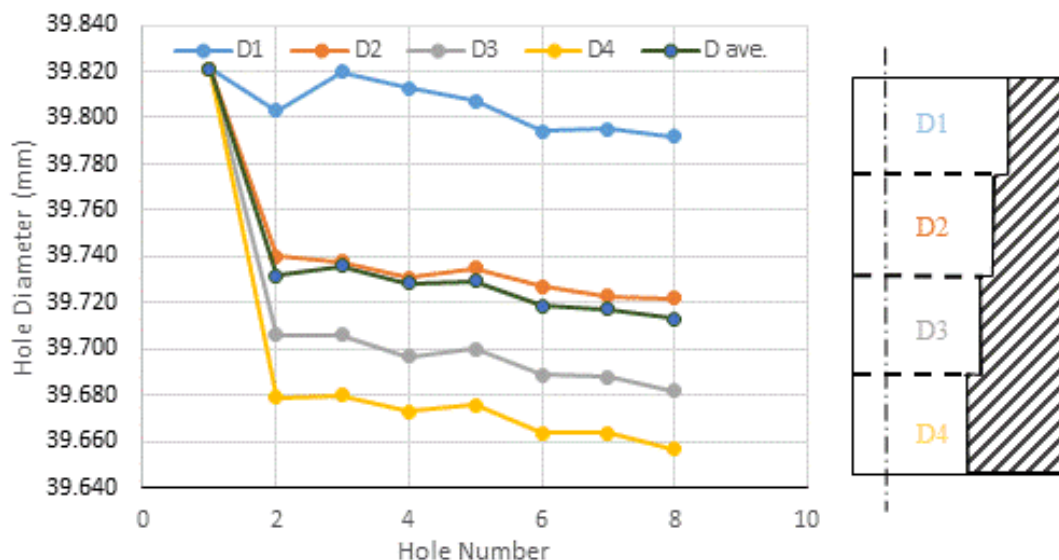


Figure 5. Variation in Cylinder Diameter

Based on the metal removal process outlined in figure 3 it must be the case that the bottom section of the cutter C4 was cutting for longer and removing more metal than the upper regions. Moving up the tool each section of tool actually removes less metal. It is also clear, from figure 5, that the diameter of the cylinders at the same depth of the corresponding regions became smaller as the milling experiments went on. In addition, there are variations between the values of the actual diameters at the different levels (D1, D2, D3, and D4) with the value of the calculated average diameter ($D_{ave.}$). This analysis suggests that the geometric form of a component machined using the uneven wear cutter will be less than optimal.

This effect can also be seen from the levels of tool wear shown figure 6. These values were calculated from the changes in cylinder diameter which represent the apparent tool wear at each of the corresponding levels, resulting in the dimensions shown in figure 6.

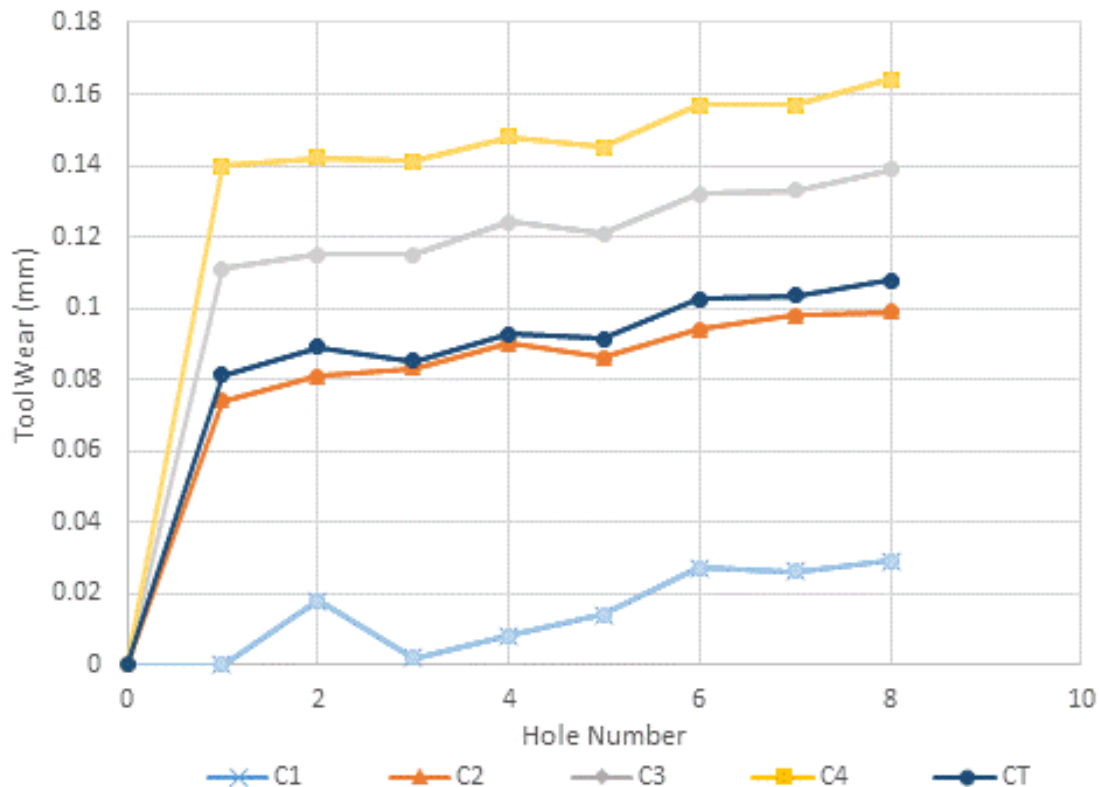


Figure 6. Tool Wear as a function of Hole Number

The occurrence of uneven tool wear is clearly shown in figure 6. As anticipated, the amount of indicated tool wear is least in the top section of the cutter (segment C1) where the maximum value reaches 0.03mm. The highest level of tool wear occurs in the section C4, which corresponds to the bottom section of the cutter, where the maximum approaches 0.16 mm. Figure 6 also indicates the results for the average tool wear value, CT, which reaches a maximum of 0.11 mm on figure 6. This indicates the potential for error associated with the non-allowance for differential tool wear in the cutter when using the average value to inform the resetting of the tool offset and also shows how this can possibly reduce the dimensional accuracy of the product.

To provide more definite assessment of the nature and level of this effect the cutting per cylinder process depicted in figure 4 was combined with the tool wear measurements in figure 6 to produce a plot of tool wear versus cutting time. This is shown in figure 7. It is important to stress that this figure represents the tool wear associated with this particular set of cutting processes. The rate and level of tool wear will obviously vary, in accordance with the expectations identified in the tool life equations.

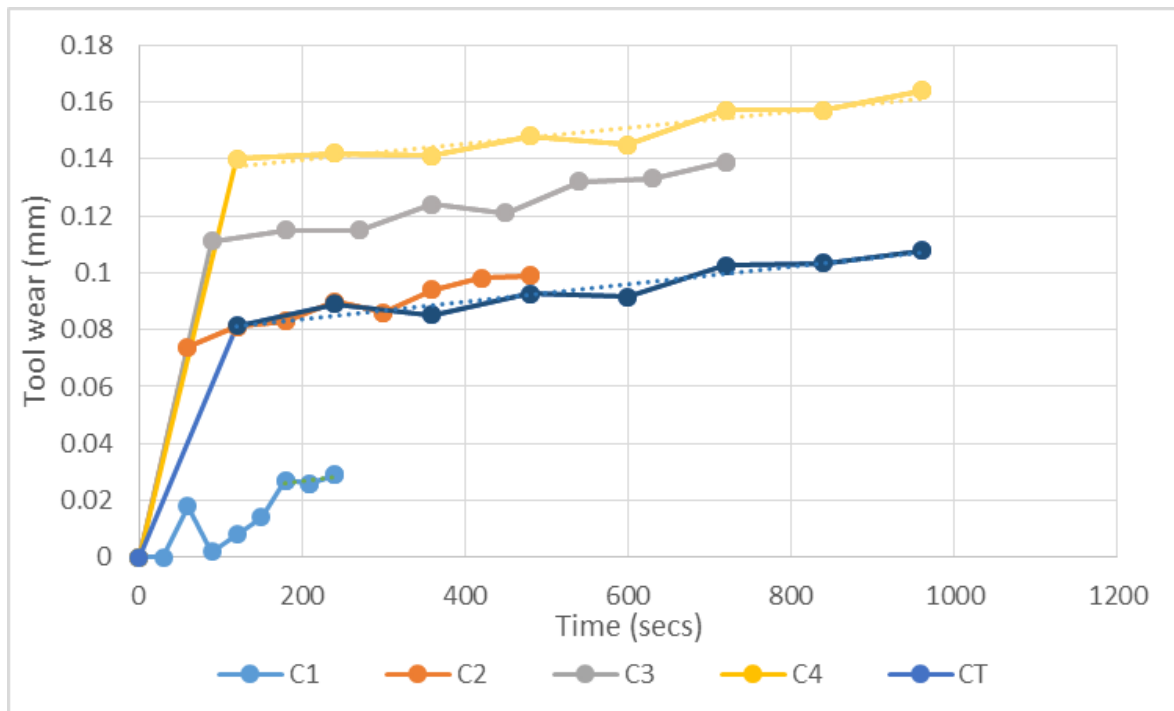


Figure 7. Tool Wear as a function of cutting time

Based upon the experimental results shown figure 7 anticipated tool life was extrapolated for three of the acquired trend lines; for C1, C4, and CT. This approach was intended to explore the potential variation in indicated tool life by estimated the cutting time at which tool wear reached the maximum advisable 0.3mm limit. From this analysis it was determined that the first segment C1 for the cutter will reach the end of usable tool life criteria at 155 minutes. The bottom section, C4, will reach to the end of tool life criteria after 92 minutes work. Finally, based on the average tool wear Ct, the cutter will reach this point after 123 minutes cutting.

The effect of basing anticipated remaining useful tool life on the measurement of tool wear is clearly important. Each of the different times indicate the error that could potentially occur. Assuming that the assessment of remaining useful life is based upon the average tool wear CT; this would indicate a value of 123 minutes. Applying a tool management strategy on this basis would mean, in this instance, the section of tool performing most of the cutting, section C4, would be at risk of failing once the cutting time passes 92 minutes. This is clearly not a viable position as it could result in tool breakage. At the other extreme, in this case, the remaining useful life of the lightly used section of cutter would be more that allowed for by using the average value; this is less important as it is not possible to make use of this section of tool and no tool failures would result.

4. CONCLUSION

This approach makes it possible to calculate the differential tool wear arising during the specified cylindrical cavity machining process. The intention of this work was to show how CMM based measurements can be used to quantify how much the different sections of the tool are worn. The analysis showed that the measurements obtained by deploying the CMM could be a reliable basis for the identification of uneven tool wear based upon the geometry of the component.

Using the accepted tool life criteria, this investigation has indicated that the flank wear arising in the bottom section of the cutter is likely to exceed the identified limit of 0.3 mm if the estimation of tool life is based upon average tool wear. In this case, the operator could continue to use a tool which is reaching a

dangerous level of tool wear because this has been hidden by the fact that the average tool wear have been taken rather than the actual value.

It is recognized that this is only a very simple investigation but it has confirmed the presence of an important process and the conclusion must therefore be that more work is required, which is the aim of this on-going research. This will support these assessments but based upon more direct tool wear-related measurements acquired from the machine tool controller.

REFERENCES

- [1] ISO 8688-2: 2011. Tool life testing in milling — Part 2: End milling. [Online]. Available at: <https://www.iso.org/obp/ui/#iso:std:iso:8688:-2:ed-1:v1:en> [Accessed: 11th February 2016].
- [2] F. W. Taylor (1906). On the art of Cutting Metal.
- [3] B. Li. (2012). A review of tool wear estimation using theoretical analysis and numerical simulation technologies. *International Journal of Refractory Metals and Hard Materials* 35, pp. 143-151.
- [4] Xiaoli, L., & Zhejun, Y. 1998. Tool wear monitoring with wavelet packet transform – fuzzy clustering method. *Wear*, 219, 145-154.
- [5] Shao H., Wang H. and Zhao X. (2004). A cutting power model for tool wear monitoring in milling *International Journal of Machine Tools and Manufacture* 44, pp. 1503–1509.
- [6] Yesilyurt I. and Ozturk H. (2006) Tool Condition Monitoring in Milling Using Vibration Analysis *International Journal of Production Research* 45, 1013-1028.
- [7] Axinte D and Gindy N (2004). Assessment of the Effectiveness of a Spindle Power Signal for Tool Condition Monitoring in Machining Processes *International Journal of Production Research* 42, 2679-2691.
- [8] Ahmed Z. J. Prickett P. W. and Grosvenor R. I. (2016). The difficulties of the assessment of tool life in CNC milling. IEEE Xplore - 2016 International Conference for Students on Applied Engineering (ISCAE)
- [9] Liu X., Zhang C., Fang J. and Guo S. (2010). A New Method of Tool Wear Measurement *2010 IEEE International Conference on Electrical and Control Engineering*.
- [10] Zhang C. and Zhou L. (2013). Modelling of Tool Wear for Ball Milling Cutter Based on Shape Mapping *International Journal Interact Design and Manufacturing* 7, 171-181.
- [11] Qin X., Zhao Q., Ji C., Li Y., Sun D., and Jin Y. (2015). Tool Life and Hole Surface Integrity Studies for Hole-making of Ti6Al4V alloy. *International Journal of Advanced Manufacturing Technology*, 79(5-8), pp. 1017-1026.
- [12] Sultan A. Z., Sharif S., and Kurniawan D. (2015). Effect of Machining Parameters on Tool Wear and Hole Quality of AISI 316L Stainless Steel in Conventional Drilling. 2nd International Materials, Industrial, and Manufacturing Engineering Conference, MIMEC2015. *Procedia Manufacturing* 2, pp. 202-207.
- [13] Kivak T., Habali K., and Seker U. 2012. The Effect of Cutting Parameters on the Hole Quality and Tool Wear During The Drilling of Inconel 718. *Gazi University Journal of Science* 25(2), pp. 533-540.
- [14] Li H., He G., Qin X., Wang G. Lu C., and Gui L. (2014). Tool wear and hole quality investigation in dry helical milling of Ti-6Al-4V alloy. *International Journal of Advanced Manufacture Technology* 71, pp. 1511-1523.
- [15] Denkena B., Boehnke D., and Dege J. H. 2008. Helical milling of CFRP-titanium layer compounds. *Journal of Manufacturing Science and Technology* 1, pp. 64-69.

Potential Use Cases of LWIR Cameras for Automation and Process Monitoring

B. Eichentopf, R. Baltes, N. Fietz, K. Nienhaus

¹ Institute for Advanced Mining Technologies, RWTH Aachen University, Aachen, GERMANY

ABSTRACT

In several industrial applications, imaging methods are commonly utilised for process monitoring, process automation and process control. In the mining and heavy machinery industry, common imaging methods, e.g. optical cameras or laser scanners often reach their limits because of the adverse environmental conditions. In these challenging environments, thermal imaging cameras based on long-wavelength infrared radiation (LWIR cameras) offer an advantage over the mentioned conventional systems [1] [2].

The main advantage of thermal imaging based on LWIR cameras is its insensitivity towards smaller particles like dust or water aerosols in the surrounding atmosphere as well as its robustness towards mechanical stress, such as vibrations. Furthermore, thermal imaging is a passive imaging method and is therefore independent of ambient light and additional, external illumination. These advantages qualify thermal imaging based on LWIR for several applications in process automation and process monitoring in the mining sector [3].

The Institute for Advanced Mining Technologies (AMT) at RWTH Aachen University has been utilising LWIR cameras for process automation and process control in different industry related projects throughout the last years. Based on shrewd image processing and determination of characteristic values, we developed different methods for the analysis of various kinds of mass flows. This enabled for example the automated feeding of water to a mass flow or the automated separation of hot spots from a mass flow. Another application realised was the automation of the feeding of a conveyor belt system through the positioning of transfer booms and transfer chutes based on LWIR images. [2]

Keywords: thermal imaging, infrared, automation, process monitoring, image processing

Corresponding author: B. Eichentopf (email: beichentopf@amt.rwth-aachen.de)

1. INTRODUCTION

The increasing level of technology and automation enabled the mining of deeper and more complex ore deposits over the last decades. Additionally, the haulage of higher amounts of bulk material as well as selective mining methods became increasingly economic. To ensure the supply with raw materials and energy economically and ecologically sustainable the development of new technical solutions and processes remains necessary. [1] [3] The focus of research on sensors and automation in mining environments has increased the availability of additional information, which can be of valuable importance for the people working in mines, e.g. through enhancing the automation processes and therefore keeping the personnel out of hazardous areas. The equipment is often exposed to harsh environments, which can be dominated by aerosols (dust and water), extreme temperatures, increased vibrations and shock, but also adverse lighting conditions. Therefore, technologies already standardised for industrial applications need to be adjusted to be utilised in the mining industry [1] [2] [3].

The Institute for Advanced Mining Technologies (AMT) at RWTH Aachen has been working on utilising sensor technologies in mining applications for more than a decade now. One of the focus areas is the

utilisation of uncooled long-wave infrared (LWIR) cameras, which have been applied for process monitoring as well as process automation purposes. This paper will illustrate two projects in which long-wave infrared cameras were utilised for thermal imaging of mass flows on conveyor belts [3] [2].

2. FUNDAMENTALS OF THERMAL IMAGING

Due to the rapid development of thermal imaging within the last decades the development of microsystems, infrared detectors as well as processing units evolved tremendously. Thermal imaging visualises the infrared radiation of objects, which is invisible for the human perception. This way, additional information, which cannot be obtained by optical cameras, can be generated and utilised for the interpretation and determination of complex processes. [3] [4]

Every object in a room with a temperature of above absolute zero (0 K) emits energy in the form of electromagnetic radiation. For ambient temperatures the maximum of the emitted radiation is allocated in the electromagnetic spectrum, with a wavelength ranging from 8 μm to 14 μm . The emitted radiation from the object is depending on the object's temperature, the surface properties as well as the direction of radiation. [3] [6] [8] Infrared cameras detect the thermal radiation of objects and visualise their information contents into pictures. Usually, they utilise the atmospheric window for visualisation, a spectral range in which thermal radiation is hardly inhibited by the atmosphere. This phenomenon is depicted in Figure 1: short-wavelength radiation is not able to pass a particle with diameter d , thus reflects on the surface of the particle. LWIR radiation on the other hand is capable of passing the particle with the same diameter and therefore has no problems with reflection. Thus, the atmospheric window is specifically suitable for passive infrared measurements. Therefore, the longer wavelength of LWIR results in higher insensitivities towards aerosols (dust and water), which make LWIR cameras most suitable for automation purposes. [2] [3] [5] [7].

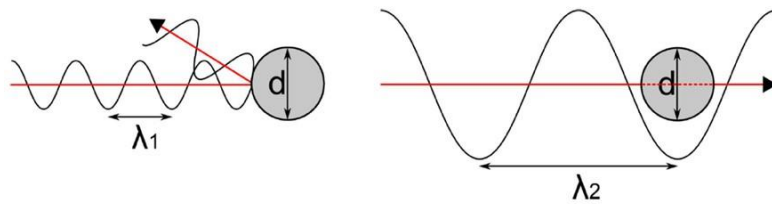


Figure 1: Short-wavelength radiation reflected by a particle (left), long-wavelength radiation passes it (right) [1]

3. APPLICATIONS

3.1. Material flow analysis

The mass flow analysis can be of crucial importance for a number of applications and can be used for the determination of the mass as well as its material composition or specific properties. According to their different emitted radiation of different sorts of material and/or their specific states it is possible to either determine the blending behaviour of different materials or their corresponding material composition. [2] [3]

The AMT conducted an industrial project in which the moisture content of the mass flow on a conveyor belt was determined. A schematic sketch is depicted in figure 2. [2] [3]

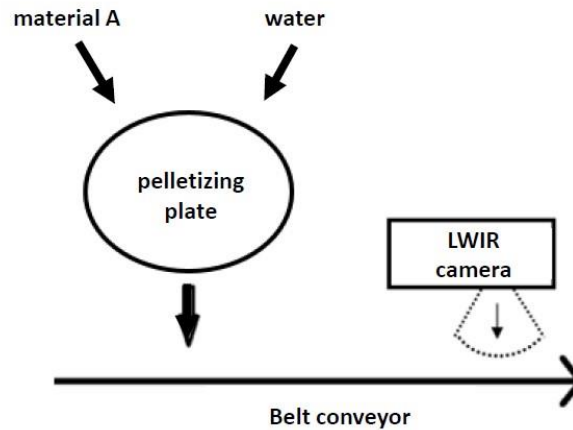


Figure 2: Schematic sketch of the blending application [3]

In the described project, the blending of material A and water needed to be monitored in order to achieve a target moisture of material A. As depicted in Figure 2 the blending was conducted on a pelletizing plate. For the given application the LWIR camera was mounted shortly after the blending process, facing the conveyor belt, so that the moisture content of material A could be determined directly after the blending process. Figure 3 illustrates two different states after blending, which were to be avoided: (left) the formation of a too moist product and (right) the formation of a too dry product [3].

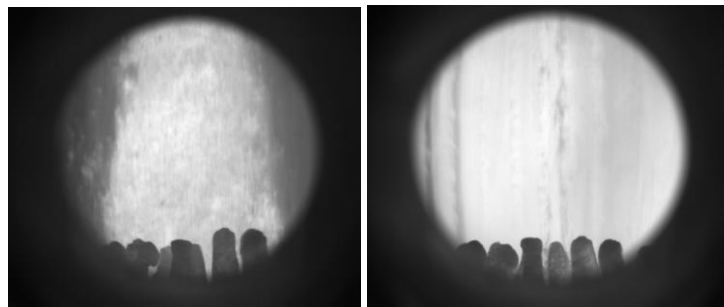


Figure 3: Two different states after blending (left: too moist product, right: too dry product) [3]

The evaluation of the generated thermal images is aided by image processing algorithms, which are capable to determine characteristic values. In combination these values form a vector in a multi-dimensional room, which describes the characteristics of the material flow. The determination of the neighbourhood assigns the vector with a specific probability to a pre-defined blending quality [3].

Figure 4 depicts the classification result for the given application according to the used processing algorithms. The x-axis describes the predicted allocation of the product conducted by the algorithm, whereas the y-axis represents the actual blending quality of the product. The green areas represent the samples, which were classified correctly and the red areas represent the samples, which were classified wrongly. In the given application the algorithm classified the too dry material correctly with a probability of 92 %. In 8 % of the cases the algorithm classified too dry material wrongly as ideal, in 0 % wrongly as too moist. Material with ideal moisture content was classified correctly as ideal in 77 % of all cases. With a probability of 6 % it was wrongly classified as too dry, with a probability of 17 % as too moist. For classification of the too moist material Figure 4 can be interpreted in the same way. Based on these results it is possible to realise the automated water inflow within the blending process. Therefore, a PID controller was implemented, which controls the water feeding process automatically. This way it was possible to reduce the needed staff for controlling the blending process [3].

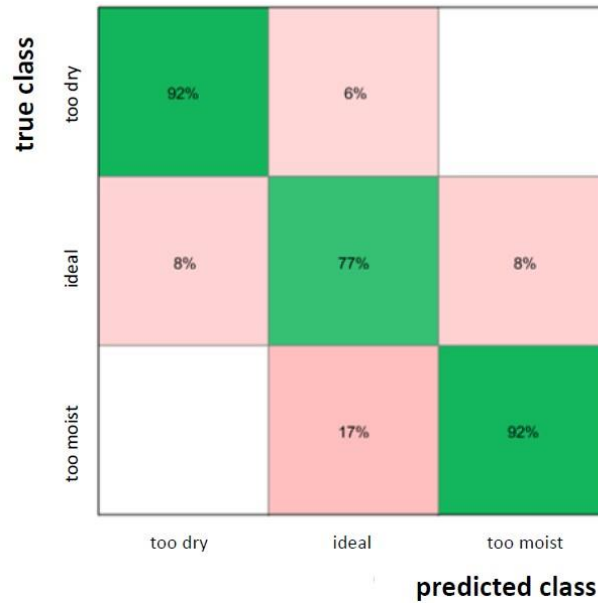


Figure 4: Classification result of the algorithm for the blending process [3]

Thermal imaging can be also utilised for the material flow analysis of bulk materials, which have areas of much higher temperature emission compared to the average mass. These areas can be of high danger for the overall production process and for the personnel working in the surrounding areas since they can result in a fire outbreak. Improved and customised image processing algorithms can detect these areas according to the increased electromagnetic radiation [3].

3.2. Belt skewing and feeder automation

Operators of open pit mines often use belt conveyors for the continuous transport of the mined material. Therefore, the total length of all belt conveyors in only one single open pit mine can easily reach several kilometres. Due to their position directly behind a bucket wheel excavator belt conveyors are of fundamental importance for the operation in open pit lignite mines. Consequently, downtimes of the belt conveyors lead to the overall interruption of the production process, because the extracted material cannot be dispatched. To minimize those downtimes the AMT developed a monitoring system for belt conveyors based on LWIR cameras in an industrially financed project. It facilitates an increase in availability and a decrease of repairing costs [2] [3].

Figure 5 schematically shows the loading process. The loading unit on the hopper car feeds the belt conveyor with material. The direction of transport of the belt conveyor in Figure 5 is assumed to be from bottom to top. The position of the hopper car can be adjusted in all three dimensions. Therefore, it is variable along the belt conveyor as well as in its vertical and horizontal angle to the belt conveyor. This flexible positioning enables an optimal loading position. In case of off-centred loading the belt can start to drift from its central position. This process is called skewing. Skewing can lead to an increase in wear and therefore to increased downtimes and costs [6] [2] [3].

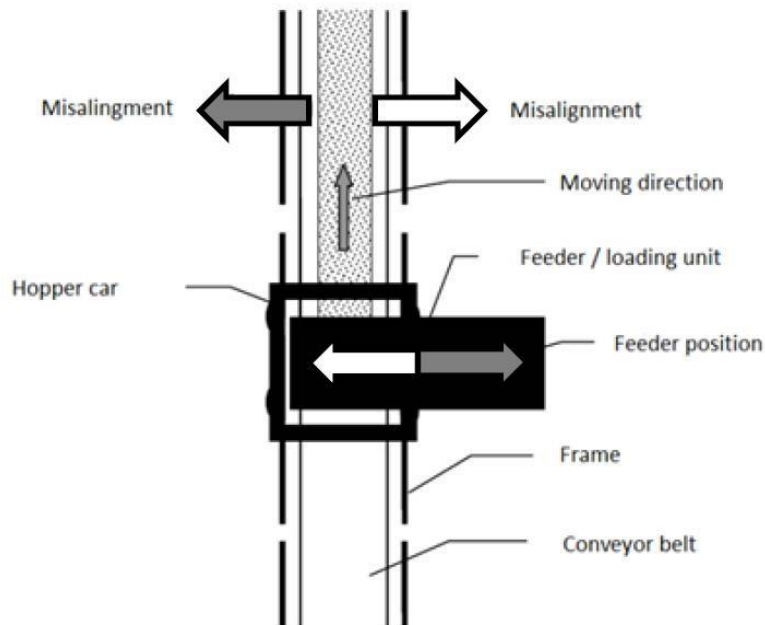


Figure 5: Schematic sketch of loading unit and hopper car [6]

To monitor the belt and detect skewing the AMT applied three LWIR cameras next to the feeder, monitoring the loading process, the incoming belt and the outgoing belt. Figure 6 depicts a thermal image of the loading process. AMT developed image processing algorithms, which compute the centre of mass of the feeding material. This information can be used to optimise the loading position [6] [3].



Figure 6: Thermal image of the loading process [6]

Figure 7 illustrates an example of a thermal image of the outgoing belt. The developed algorithms relate the detected edges to their position in the thermal image. Therefore, potential belt skewing can be identified and the loading process optimised. The result can either be used for process control or visualised for the machine operator [6] [3].

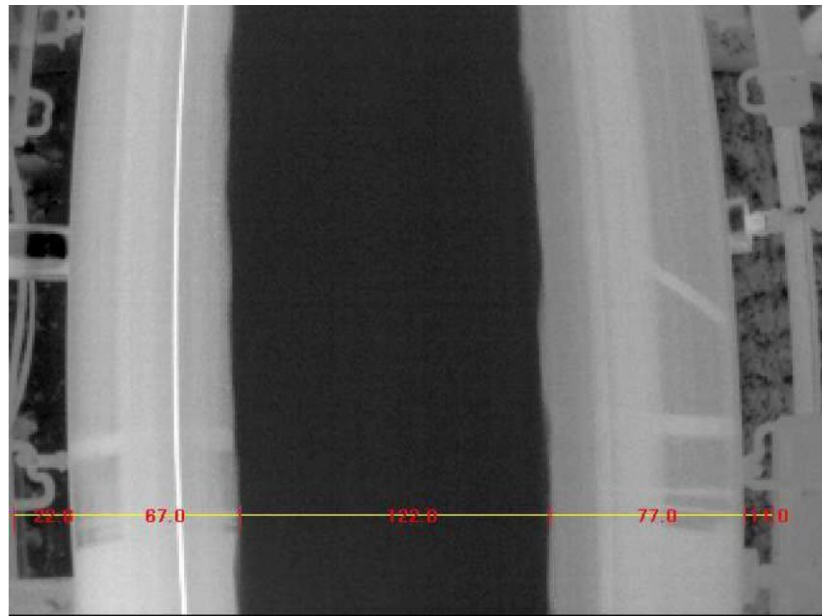


Figure 7: Thermal image of the outgoing belt [6]

4. CONCLUSION

Thermal imaging is well suited for different applications in the mining and heavy machinery industry with respect to automation, monitoring and process control. Due to the used wavelength, LWIR cameras are less sensitive towards aerosols in the ambient air. Using various image processing algorithms the AMT extracts the necessary information from the thermal images. For belt skewing and feeder automation geometrical information are processed to compute the centre of mass of the loaded material and to detect belt skewing. For the described moisture content analysis of the mass flow geometrical information and differences in the radiation intensity are combined to control the water addition [3].

Especially adding the information of the exact temperatures of the object into the image processing will open new research fields for thermal imaging in process and machine monitoring. Therefore, automation solutions such as based on thermal imaging will probably become more common in industrial applications [2] [3].

REFERENCES

- [1] Berg, J.; Baltes, R.: Einsatzmöglichkeiten von langwelligen Infrarotkameras in rauen Umgebungsbedingungen. ThermographieKolloquium 2015.
- [2] Bartnitzki, T.: Forschung für den Bergbau 4.0. Bergbau, Energie und Rohstoffe 2015. pp. 234 – 240
- [3] Eichentopf, B.;
- [4] Vollmer, M.; Möllmann, K.: Infrared Thermal Imaging. Fundamentals, Research and Applications. Wiley-VCH, Weinheim, 2013.
- [5] Nienhaus, K.; Pretz, T.; Wotrub, H.: Sensor Technologies: Impulses for the Raw Materials Industry. Shaker, Aachen, 2014.
- [6] Nienhaus, K.; Warcholik, M.; Büschgens, C.; Müller, D.: Belt Positioning and Skewing Prevention in Lignite Mining Using LongWavelength Infrared Cameras. 12th International Symposium on Continuous Surface Mining (ISCSM), Aachen, 2014.
- [7] Fouad, N. A.; Richter, T.: Leitfaden Thermographie im Bauwesen. Theorie, Anwendungsgebiete, praktische Umsetzung. Fraunhofer, Stuttgart, 2008.
- [8] Schuster, N.; Kolobrodov, V.G.: Infrarotthermographie. Wiley-VCH, Weinheim, 2004.

Session 9

Monitoring, Diagnosis, Prognosis and Health Management 3

Session Chair Dr. Hao Zhang

Condition Monitoring of Guiding Systems by Means of Model-Based Virtual Sensors

(M. Gonzalez, O. Salgado, J. Croes, B. Pluymers, W. Desmet)

A Kernel Fisher Discriminant Framework Based on Kernel Entropy - Component Analysis for Fault Diagnosis of Rolling Bearings *(H. Zhou, T. Shi, W. Lai, G. Liao, J. Xuan)*

Drill Bit Wear Monitoring Based on Vibration Signal Analysis *(H. Rafezi, F. Hassani)*

Detection and Diagnosis of Reciprocating Compressor Faults Based on Modulation Signal Bispectrum Analysis of Vibrations *(U. Haba, K. Brethee, O. Hassin, F. Gu, A. Ball)*

Condition Monitoring of Guiding Systems by Means of Model-Based Virtual Sensors

M. Gonzalez^{1,2}, O. Salgado¹, J. Croes^{2,3}, B. Pluymers^{2,3}, W. Desmet^{2,3}

¹ IK4-Ikerlan Technology Research Center, Control and Monitoring Area, Spain

² KU Leuven, Department of Mechanical Engineering, Belgium

³ Member of Flanders Make, Belgium

ABSTRACT

Condition monitoring schemes rely on the measurement of specific variables of the system. In many applications, however, the required variables are either too costly or impossible to measure. Model-based virtual sensors are an attractive option to overcome this challenge. In this paper a practical application of this approach is presented. Model-based virtual sensors are used to estimate the forces acting on the guiding rails of a vertical transportation system. The friction coefficient is estimated along with the contact forces. The effect of different defects on the estimated features is studied. This is the first step to define a criterion for the detection of non conformities in guiding systems by means of model-based virtual sensors.

Keywords: Model-based condition monitoring, Virtual sensing, Kalman filtering

Corresponding author: O. Salgado (email: osalgado@ikerlan.es)

1. INTRODUCTION

Maintenance services account for a significant part of the operating cost of high-value systems, especially when possible failures result in large down times [1]. Condition based maintenance strategies provide a cost efficient way to cope with the safety and reliability requirements of such systems [2]. These strategies monitor the condition of the system to evaluate the maintenance requirements. Condition monitoring schemes rely on the measurement of specific variables of the system. These measurements are later processed with fault detection methods in order to extract features that allow us to detect faulty conditions of the asset.

A review of fault and damage detection methods can be found in [3], [4] and [5]. These methods may be divided in three main categories: data-driven methods, hybrid approaches and model-based methods. Data-driven methods use a machine learning perspective to recognize patterns in the behavior of the system. These methods require previous observations of the system in order to learn the different states and conditions of the asset. Such “training” can be a handicap in systems where data cannot be acquired continuously or in which faulty conditions cannot be measured. Hybrid methods are aimed at solving these difficulties [6]. Hybrid methods complement missing data of the system with artificial data simulated using a first principles model of the system. Model-based methods combine first principles models and measurements of the system by means of estimation algorithms. The model provides knowledge of the dynamics of the system, which in combination with off-the-shelf sensors is used to estimate variables of interest otherwise difficult to measure. This is known as a virtual sensor, i.e. a measure of an unknown variable of the system that we are not measuring directly, but which we derive from the estimated states and parameters instead. Model-based approaches are a valuable tool in applications where the direct measurement of the variables of interest is not feasible. The efficacy of these techniques highly depends on the accuracy of the model [7]. This can be a major drawback of these techniques, as the high complexity of modern electromechanical systems results in models which are either too costly or not accurate enough. Therefore specialized tools are required to cope with the complexity of the models.

This paper deals with the application of a condition monitoring method for the identification of damages in guiding rails. The condition of the guiding system influences significantly the riding quality and performance of transportation systems such as railways or elevators. The proper design and the correct maintenance of the guides is therefore of high importance. The guiding system is not the most common source of failure but it is one of the most critical systems. Failure in guiding rails leads to large down times of the whole system and it is difficult to evaluate its condition. This is why mayor players in the lift and train industries are investing in the integration of condition monitoring system with their new smart systems.

Different condition monitoring alternatives are suggested in the literature for guiding rails. Commonly data-driven methods are used, which require specific ad-hoc sensors and well trained data. This is a handicap of these methods, as ad-hoc sensors are commonly too intrusive and complicated to install. Additionally data driven methods require exhaustive training, which in many cases is difficult to obtain. In particular, guided systems are generally customized products (i.e. every installation will have different characteristics) requiring a new training for each new system and making a training with damages unfeasible.

For instance, in [8] a method for processing images of the surface of the guiding rails to assess their condition is proposed. This method allows assessing wear in the guiding rails but it requires a complex set of cameras to be mounted on the elevators cage roof. In [9] vibration signals of an elevator are processed with optimal wavelet packets and a support vector machine classifier to construct fault features. This SVM requires well-trained data to properly estimate the mentioned features, which in many occasions may not be available. In [10] wheel impact load detectors are used to monitor de condition of railway vehicles, installing such a device in vertical transportation system is however unfeasible. The work in [11] and [12] uses acceleration measurements to track changes in the dynamic behavior of the rail-wheel interaction. They later relate these changes to faults in the rail joints, such as bolt tightness, cracks in the fasteners and damaged insulation layers. In [13] accelerometers mounted on the train are used to track irregularities. They use a mathematical model of the train to calculate the relation between the displacement of the guide rails and the position of the sensors. After integrating twice the accelerations to get displacement of the train and they use the transfer function between the track irregularities and the computed displacements to get the displacement of the guides. However the transfer function they use is a linear one, neglecting thus relevant nonlinearities. In the field of model-based methods in [14] and [15] a model of the train is used with a Kalman filter and least squares respectively to estimate the forces in the guides and use these estimations to assess the condition of the guides. In [16] model-based approaches are used to estimate friction forces in vertical transportation systems, however lateral dynamics of the system are neglected, making it difficult to assess the condition of the guiding system.

Monitoring of guiding systems requires an accurate assessment of the loading condition. However the direct measurement of forces is not feasible, as a dedicated sensor is too costly and intrusive. Model-based virtual sensors are an attractive option to overcome these difficulties. In contrast to other fault and damage detection methods, model-based techniques incorporate physical knowledge of the system, allowing a deeper understanding of the process behavior [7]. Therefore these techniques provide not only crucial information on unobservable quantities but also physical insight into why the system's performance is degrading. Furthermore, reusing the model in different systems is possible by just changing some parameters, instead of requiring new training for each system.

The rest of the paper is organized as follows. Section 0 gives an overview of Model-based virtual sensors and explains the algorithms used in the current application. Section 0 describes the proposed application case along with the proposed model, a test bench of a vertical transportation system where contact and friction forces are estimated. The results of these estimations are shown in section 0. The final conclusions and the future work are drawn in section 0.

2. MODEL-BASED VIRTUAL SENSOR APPROACHES

The core of model-based virtual sensors consists on the use of state estimation algorithms. These algorithms use the difference between the real measurements and the prediction of a physics-based model to correct the output of the model. The presented work is focused on Bayesian estimators, particularly in the Kalman Filter based algorithms, as these are the most commonly used estimators.

The Kalman Filter (KF) is the optimal linear estimation filter [17] and the most widely used Bayesian estimator. The most common derivation of the KF uses a linear model as the one defined in equation (1). As the model is linear, the matrices \mathbf{F} , \mathbf{G} and \mathbf{H} are constant. Additionally, both the process and measurement equations are assumed to be disturbed by zero mean white Gaussian noise (\mathbf{w} , \mathbf{v}) of covariance \mathbf{Q} and \mathbf{R} respectively. The states are assumed to be Gaussian variables with a covariance \mathbf{P} and mean the state estimation ($\hat{\mathbf{x}} \sim \mathcal{N}(\hat{\mathbf{x}}, \mathbf{P})$).

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \end{aligned} \quad (1)$$

The KF algorithm is shown in figure 1. In each k-time step the system model is evaluated and compared against measured data. This is done in two steps: prediction and update. In the prediction step an a-priori estimation of the states mean and covariance is obtained from the system's model. In the update step this a priori estimation is corrected using the system's output. This estimation process is done recursively: all the prior information is summarized in the initial mean and covariance of each step ($\hat{\mathbf{x}}_k^+, \mathbf{P}_k^+$). Therefore the computational effort in each time step is the same regardless the number of measurements.

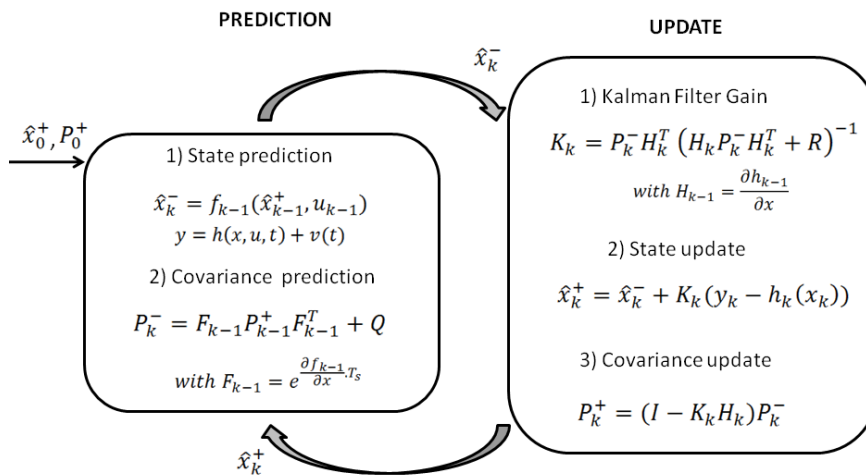


Figure 1 Kalman Filter Algorithm

Despite being widely used, the KF is limited to linear systems, which also makes the joint estimation of states and parameters not applicable [17]. Several suitable extensions of the KF to non-linear systems can be found in [17]. As the current system is highly nonlinear (events, rotations) the well known Extended Kalman Filter is used instead.

The EKF is the most widely used extension of the KF for nonlinear systems and for the joint estimation of states and parameters. If the model of equation (1) is nonlinear the \mathbf{F}_{k-1} and \mathbf{H}_{k-1} matrices of equation (1) are no longer constant, but change at each K-step instead. The EKF linearizes and discretizes the model around the KF estimate, propagating a linear approximation of the covariance [17]. The standard KF shown in figure 1 is then applied at this linearized point. As the estimation is based on the linearization of the system a small step size is required if the system is highly nonlinear. On the other hand, the ease of implementation and the reduced computational cost of the EKF make it an attractive option for the estimation of nonlinear systems.

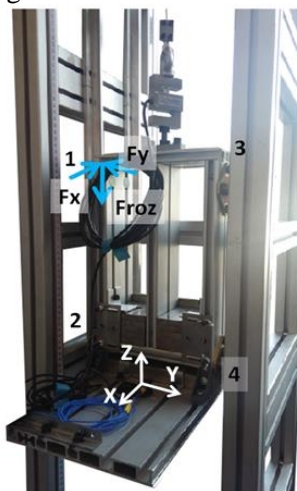
State estimation algorithms can be augmented to estimate not only the states of the system but unknown parameters too. Based on the continuous state-space system representation, an augmented version of the system can be obtained if the unknown parameters are included in the states vector and their directional derivatives are included in the system matrices. Then a random walk model is used for the unknown parameters: they are assumed to remain constant except for an additive noise [18].

Once all the states and parameters of the model are known, we can use the model to obtain some other variables of interest (i.e., a virtual sensor). This is a post-processing step in which the model is evaluated in the estimated set of states, parameters and inputs and the variables of interest are treated as another model output. By means of the estimated state covariance the degree of uncertainty of the virtual sensors can be estimated as well.

3. APPLICATION CASE AND MODEL

The proposed application case for the condition monitoring of guiding rails is the guiding system of a vertical transportation system. A scaled test bench of a vertical transportation system available at IK4-Ikerlan is used to evaluate the performance of the proposed methodology. This test bench is a useful tool to study the behavior of such system's using sensors not available in real installations. Additionally it allows us to study the effect of defects that we could not study in a real installation.

The test bench is a scaled 'rucksack' type vertical transportation system. In this paper only the dynamics of the rigid car frame are taken into account. The car travels in vertical direction and is constrained horizontally by two T-shaped guiding rails (see figure 2). The system has 12 states corresponding to the 6 degrees of freedom of this car frame ($x, y, z, roll, pitch, and yaw$). Without loss of generality, the driving force (T) is assumed to be known and acts as a system input. In the scaled test bench under study this force is measured with a load cell attached to the driving cable (see figure 2). In a real installation forces can be measured using off-the-shelf sensors such as the *Henning weightwatcher*. The T-shaped guiding rails are used to minimize horizontal motion ensuring travel in a uniform vertical direction. Inappropriate installation of these guides and their surface roughness are the main causes of vibration in the car frame [19]. These guides are usually composed of several rail segments aligned together. The proper alignment is, however, extremely difficult, and in general out of plane or out of angle misalignments are common. Small deviations increase the contact forces and induce abrupt forces in the car frame at the rail segment joints, reducing the ride quality and efficiency of the system. The numbering followed in this paper for the contact forces is shown in figure 2.



| PARAMETER | DESCRIPTION | UNITS | VALUE |
|--------------------------|------------------------|----------------------|----------------------------|
| M | carframe's mass | [Kg] | 14.287 |
| I_{11}, I_{22}, I_{33} | Carframe's inertias | [Kg.m ²] | 0.28, 0.38, 0.20 |
| K_x, K_y | contact stiffness | [N/m] | 600000 |
| D_x, D_y | contact damping | [N/(m/s)] | 10 |
| ClearanceY | sliding shoe clearance | [m] | 0.0 |
| r_1^0 | position of shoe 1 | [m] | (-0.085, -0.124, 0.297) |
| r_2^0 | position of shoe 2 | [m] | (-0.085, -0.124, 0.067) |
| r_3^0 | position of shoe 3 | [m] | (-0.085, 0.124, 0.297) |
| r_4^0 | position of shoe 4 | [m] | (-0.085, 0.124, 0.067) |
| $r_{c.g}^0$ | carframe's C.G | [m] | (-0.0923, 0.0043, 0.08824) |
| r_T^0 | position of cable | [m] | (-0.085, 0.0, 0.435) |

Figure 2 Test bench of the application and properties

The high complexity of modern electromechanical systems requires specialized tools for the development of the models. In addition a great modeling flexibility and simplicity is required to avoid errors and speed up the modeling process. Thus the modeling language Modelica is used to model the system [20]. The

acausal nature of Modelica allows efficiently modeling heterogeneous systems reusing already developed and tested models. However, it doesn't allow the user to manipulate the solution at each time step, as required by estimation algorithms. The Functional Mock-up Interface (FMI) is a tool independent standard that can efficiently solve this [21]. The combination of Modelica with other programming languages by means of the FMI provides thus a suitable approach for the implementation of model-based virtual sensors. Therefore these tools are used for the development of the model of the current system.

The inertial properties of the cabin given in figure 2 are obtained from its CAD model. The contact stiffness can be obtained from classical structural analysis, assuming the guide as a flexible beam with flexible supports. In order to simplify the estimation we assume a constant stiffness for the whole guide. Due to the low contact forces and the high stiffness of the car frame, the latest can be modeled as a rigid body. The movement of the body is represented in the coordinate system attached to the floor of the car frame as it is the location where the required sensors are installed. The rotation is constraint far away from the Gimbal lock position due to the guide rails. Hence rotations of the car frame are represented using Euler angles (α , β and γ). The sliding shoes are the interface between the car frame and the guiding rails. From figure 3 it can be seen that movement of the shoe in the x direction will always result in a force opposite to the movement. Thus, for modeling purposes, contact in x axis can be assumed to behave as a spring-damper system. Movement of the sliding shoe in the y axis will depend on the direction of movement. Hence, contact in Y direction is modeled using an event driven spring damper. Coulomb friction is used to model friction.

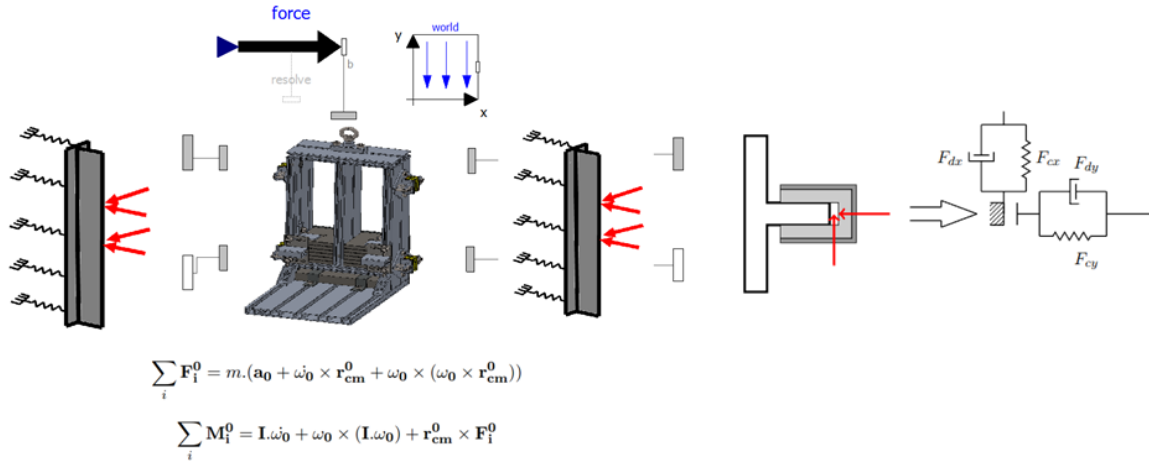


Figure 3 Model of the system in Modelica, with the cabin modeled as a rigid body and the contact with the guides as spring-dampers.

4. RESULTS

In this section the results of the application of model-based virtual sensors for the estimation of forces in guiding rails is presented. The estimation approach from section 0 is applied to the system described in section 0. The measurements used for the EKF are the lateral and vertical accelerations and the vertical position of the car. These measurements are taken with a triaxial piezoelectric accelerometer (lateral acceleration), with a DC response accelerometer (vertical acceleration) and with a draw-wire encoder (cabin position). The measurement of the lateral acceleration provides information of the dynamic change of the contact forces, induced by the roughness and defects of the guiding rails. In addition the model provides information regarding the dynamic behavior of the car and the order of magnitude of the forces. However, the misalignment of the guiding rail results in a DC component (0 Hz) of the contact forces which cannot be estimated, as neither the lateral acceleration nor the model have information on that regard. This misalignment affects the vertical dynamics of the system, as friction increases with it. Therefore, we can account for this effect within the friction coefficient μ . Thus this parameter is estimated jointly with the states of the system as explained in section 0. This parameter contains thus information both on the actual

coulomb coefficient and on the misalignment of the guiding system. Together with the contact forces this parameter provides a suitable indicator of the condition of the guiding rails.

Different defects are introduced in the guiding system and their effect on the estimation is assessed. The estimation of contact forces and friction coefficient for three different defects is shown in this paper. Two defects are introduced in the web of the guiding rail, one in X direction (height 1.22m) and one in Y direction (heights 0.98m and 1.48). Figure 4 shows the defects introduced in the web of the guiding rail. In the third defect the joint of the guiding rail located at height 1.22 m is moved circa 2mm from its original position to emulate a misalignment of the guide.

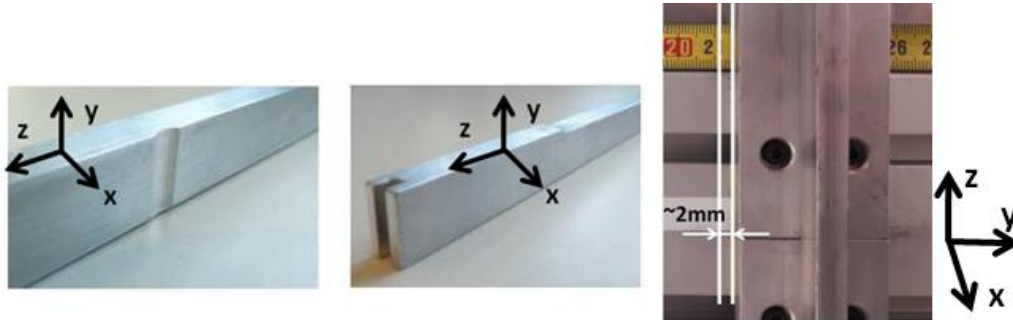


Figure 4 Defects in the guides

Figure 5, figure 6 and figure 7 show the results of the estimated contact forces and the friction coefficient. The friction coefficient has information regarding the actual friction coefficient (a function of lubrication, frequency of use and others) and the misalignment of the guides. On the other hand the estimated forces give us information about local defects on the guiding rails. The horizontal dash lines show the position of the joints of the guiding rail.

Figure 5 shows the estimation result for a circa 2mm misalignment of the guides at height 1.22m. The estimation shows that there is a significant increase of the friction coefficient at the height of the defect and a change in the contact forces with respect to the base case. The estimation also shows a relevant increase of the friction coefficient at height 0.7m, meaning that a wrong position of one of the guide joints causes a deformation of the guide at different heights. The estimation of this parameter is, as expected, directly related to the misalignment of the guides. However, further work is required to define an unbiased criterion for the location of the misaligned joint. Figure 6 shows the estimation for a defect in the Y direction of the web of the guide. The defect causes an abrupt change in the forces in Y direction, while there is no relevant change in the friction coefficient. Although the change is dominant in the forces in Y direction, the forces in X direction are also affected. This effect has to be taken in to account when defining criteria for damage detection, in order to avoid false positives. Finally figure 7 shows that the defect introduced in X direction has had no relevant effect on the system. It is probable that the defect is too small to have an effect on the system. However, even though the change is not as relevant as in the case with a damage in Y direction, forces in Y direction have increased. Therefore it is clear that the estimation of the forces is more sensitive to other effects than the estimation of the friction coefficient.

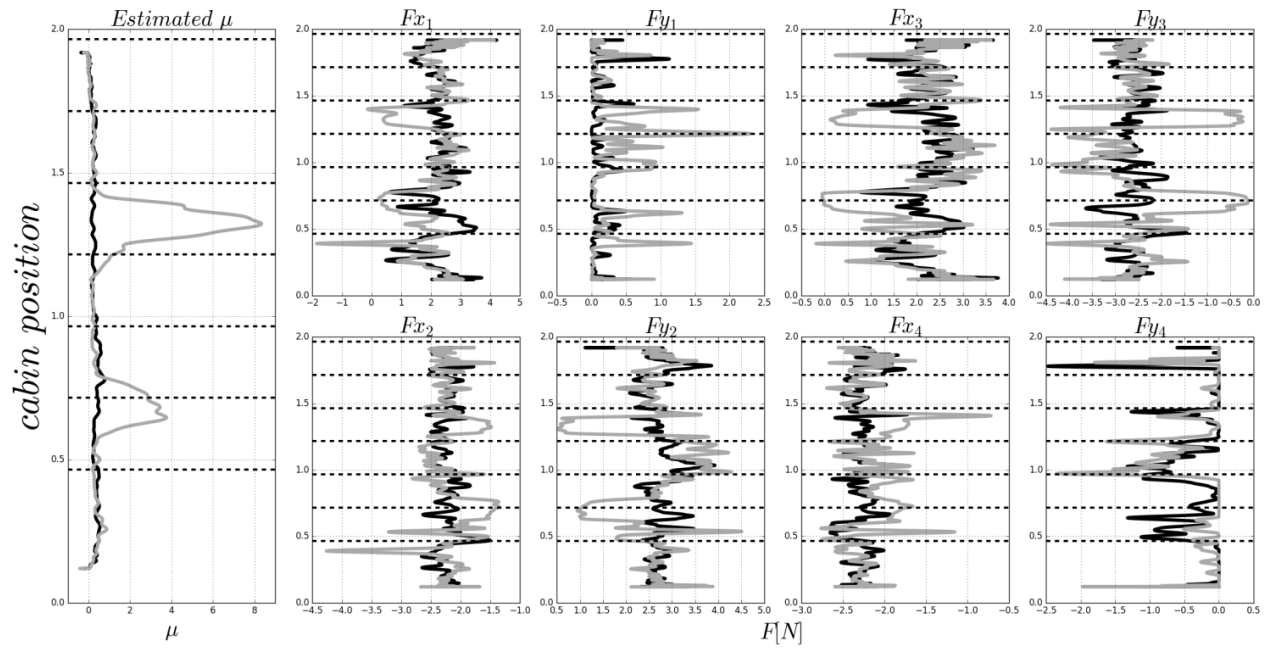


Figure 5 Estimation with a deviation of 1.5mm of the guide joint at height 1.22m. In black the base case, light grey the damaged case.

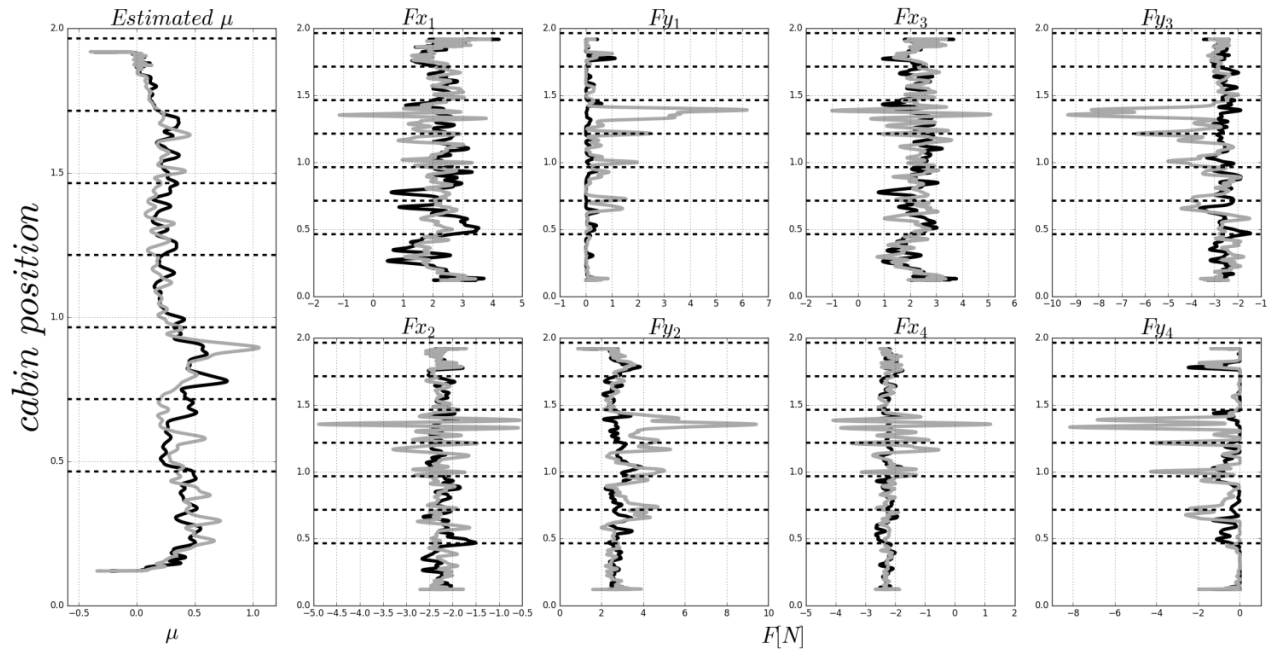


Figure 6 Estimation with a defect in Y direction of the guiding rail's web. In black the base case, light grey the damaged case.

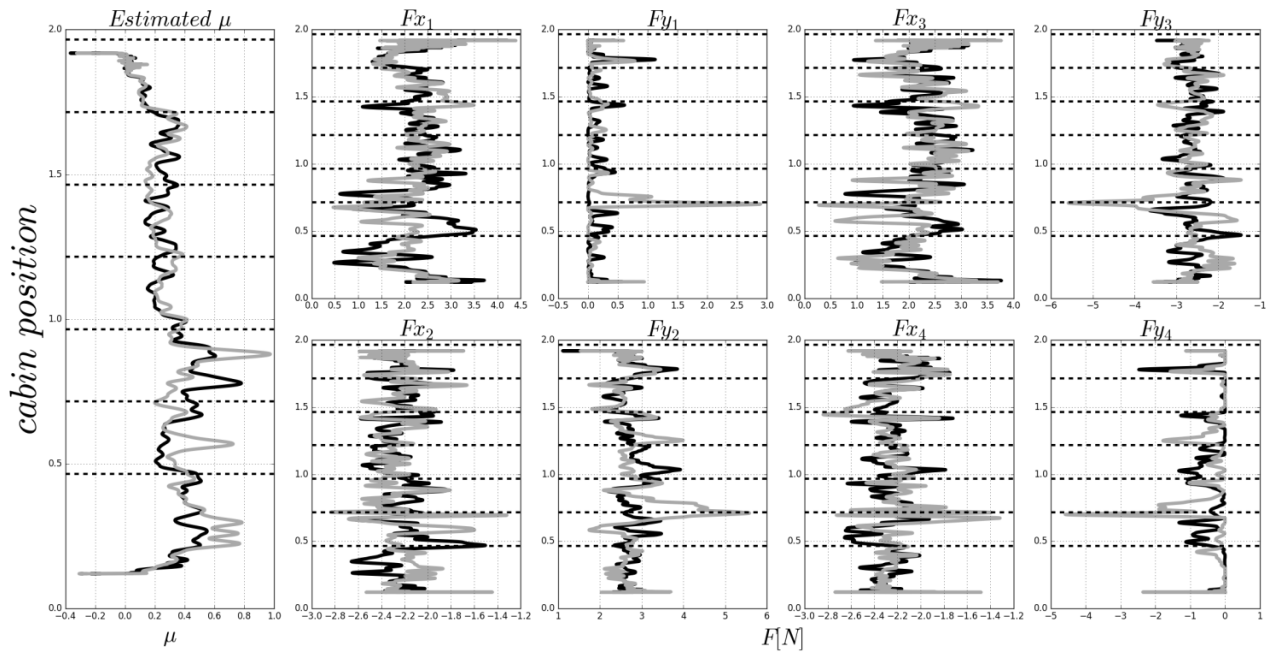


Figure 7 Estimation with a defect in X direction of the guiding rail's web. In black the base case, light grey the damaged case.

5. CONCLUSIONS AND FUTURE WORK

Model-based virtual sensor approaches have been used to estimate contact forces and the friction coefficient in the guiding rails of a vertical transportation system. These estimations are a useful feature to evaluate the condition of the guiding system. Misalignment and local defects of the guides can be detected using off-the-shelf sensors. This approach opens the way to condition based maintenance strategies for guiding systems of vertical transportation systems. Such maintenance schemes can improve riding quality, safety and efficiency of vertical transportation systems, fulfilling thus the requirements of modern smart installations. Further work is required mainly in two aspects of the investigation:

- Assessment of the estimated variables and parameters: the results of the defects are compared against one base case to show that the estimated features can be used as damage indicators. However in a general case there is no base case available. A theoretical optimal value of the friction coefficient and of the contact forces should be used to assess the condition of the guides. In other words, we need a criterion to identify damage in an unbiased way. Furthermore, the effect of other conditions such as lubrication stats of the system has to be studied. The model of the system will be used to analyze in deep the effect of the defects.
- In this work the tension in the cable has been used as input of the system. Even though off-the-shelf sensors exist for that purpose it would be better to use just sensors available in the system or easier to use, such as the input of the controller and the currents in the machine. Thus the model of the system has to be extended to include not only the cabin of the test bench but the controller, the electric machine, the pulley, the cable and the counterweight as well.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the European Commission for its support of the Marie-Sklodowska Curie program through the ITN ANTARES project (GA 606817) and the support from the KU Leuven research fund.

REFERENCES

- [1] P. Amorim-Melo, E. Shehab, L. Kirkwood, and P. Baguley, "Cost drivers of integrated maintenance in high-value systems," *Procedia CIRP*, vol.22, pp. 152–156, 2014.
- [2] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 2012.
- [3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part i: Quantitative model-based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 293–311, 2003.
- [4] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 313–326, 2003.
- [5] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part iii: Process history based methods," *Computers & chemical engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [6] U. Leturiondo, O. Salgado, L. Ciani, D. Galar, and M. Catelani, "Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data," *Measurement*, 2017.
- [7] R. Isermann, "Model-based fault-detection and diagnosis–status and applications," *Annual Reviews in control*, vol. 29, no. 1, pp. 71–85, 2005.
- [8] O. Yaman, M. Baygin, and M. Karakose, "A new approach based on image processing for detection of wear of guide-rail surface in elevator systems," *International Journal of Applied Mathematics, Electronics and Computers*, 2016.
- [9] Z. Wan, S. Yi, K. Li, R. Tao, M. Gou, X. Li, and S. Guo, "Diagnosis of elevator faults with ls-svm based on optimization by k-cv," *Journal of Electrical and Computer Engineering*, vol. 2015, p. 70, 2015.
- [10] M. Palo, H. Schunnesson, and U. Kumar, "Condition monitoring of rolling stock using wheel/rail forces," in *International Conference on Condition Monitoring and Machinery Failure Prevention Technologies: 12/06/2012-14/06/2012*, vol. 1, 2012, pp. 560–568.
- [11] M. Oregui, S. Li, A. Núñez, Z. Li, R. Carroll, and R. Dollevoet, "Monitoring bolt tightness of rail joints using axle box acceleration measurements," *Structural Control and Health Monitoring*, 2016.
- [12] M. Molodova, M. Oregui, A. Núñez, Z. Li, and R. Dollevoet, "Health condition monitoring of insulated joints based on axle box acceleration measurements," *Engineering Structures*, vol. 123, pp. 225–235, 2016.
- [13] X. Wei, F. Liu, and L. Jia, "Urban rail track condition monitoring based on in-service vehicle acceleration measurements," *Measurement*, vol. 80, pp. 217–228, 2016.
- [14] G. Charles, R. Goodall, and R. Dixon, "Model-based condition monitoring at the wheel–rail interface," *Vehicle System Dynamics*, vol. 46, no. S1, pp. 415–430, 2008.
- [15] G. Charles, R. Dixon, and R. Goodall, "Least squares method applied to rail vehicle contact condition monitoring," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 7451–7456, 2008.
- [16] E. Esteban, O. Salgado, A. Iturrospe, and I. Isasa, "Model-based estimation of elevator rail friction forces," in *Advances in Condition Monitoring of Machinery in Non-Stationary Operations*. Springer, 2016, pp. 363–374.
- [17] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.
- [18] F. Naets, J. Croes, and W. Desmet, "An online coupled state/input/parameter estimation approach for structural dynamics," *Computer Methods in Applied Mechanics and Engineering*, vol. 283, pp. 1167 – 1188, 2015.
- [19] J. Andrew and S. Kaczmarczyk, *Systems Engineering of Elevators*. Elevator World, 2011.
- [20] H. Elmqvist, S. E. Mattsson, and M. Otter, "Modelica-a language for physical system modeling, visualization and interaction," in *Computer Aided Control System Design, 1999. Proceedings of the 1999 IEEE International Symposium on*. IEEE, 1999, pp. 630–639.
- [21] T. Blochwitz, M. Otter, J. Akesson, M. Arnold, C. Clauss, H. Elmqvist, M. Friedrich, A. Junghanns, J. Mauss, D. Neumerkel et al., "Functional mockup interface 2.0: The standard for tool independent exchange of simulation models," in *Proceedings of the 9th International MODELICA Conference; September 3-5; 2012; Munich; Germany*, no. 076. Linköping University Electronic Press, 2012, pp. 173–184.

A Kernel Fisher Discriminant Framework Based on Kernel Entropy Component Analysis for Fault Diagnosis of Rolling Bearings

Hongdi Zhou¹, Tielin Shi¹, Wuxing Lai¹, Guanglan Liao¹, Jianping Xuan¹

¹ State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

ABSTRACT

Kernel entropy component analysis (KECA) is a newly developed information-theory-based dimensionality reduction and feature extraction method, which has been widely used for pattern recognition. KECA can reveal the structure related to Renyi entropy of the input data set, and don't necessarily depend on the top eigenvalues and corresponding eigenvectors of the special constructed kernel matrix. However, KECA is an unsupervised method which may discard discrimination information and weaken the classification performance. This paper presents a hybrid method which combines the KECA and the Fisher's linear discriminant analysis (LDA) for feature reduction in fault diagnosis of rolling bearings. The method can make full use of the descriptor of information entropy and discriminant information of classes to improve the classification performance. High-dimensional features based on wavelet packet entropies derived from the raw vibration signals are firstly constructed from the original vibration signals. Then the intrinsic geometric features extracted with KECA plus LDA are fed into the support vector machine classifier to identify different operating conditions of bearings. The experimental results demonstrate the feasibility and effectiveness of the proposed method.

Keywords: fault diagnosis, kernel entropy component analysis, Renyi entropy, dimension reduction, pattern recognition, feature extraction

Corresponding author: Wuxing Lai (wxlai@hust.edu.cn)

1. INTRODUCTION

Rolling element bearings (REBs), the most common parts in almost all kinds of rotating machines, play an exceptionally vital role in modern industry due to their relatively lower price, interchangeability and operational ease [1]. In industrial applications, the proper functioning of the machinery system is entirely depended on the condition of REBs, which accounts for almost 45–55% of these equipment failures [1-3]. Unexpected bearing defects (such as galling, peeling and fatigue due to misalignment, shaft slope and surface roughness, etc.) reduce the reliability and availability of the equipment, which may cause huge economic losses and even lead to casualties [2]. Thus, it is necessary to study and implement effective maintenance strategies to diagnose the incipient bearing faults in rotating machines.

Vibration-based fault diagnosis has been extensively studied to improve existing techniques toward the goal of more accurate dealing with different conditions, such as varying load effect and noise contamination. The major challenge is extracting reliable and sensitive features from the original signals to reflect the operating condition of the systems [4]. To comprehensively describe the condition, a high-dimensional feature set including time-domain features, frequency-domain features and time-frequency domain features are often extracted for diagnosis [4-6]. Undoubtedly, more features can provide more information, while they also have redundant information which would increase computation time and reduce recognition accuracy. Then effective feature extraction and dimension reduction methods are crucial for machine health diagnosis.

Moreover, advanced dimension reduction techniques, such as principal component analysis (PCA), kernel principal component analysis (KPCA) [7], Fisher's linear discriminant analysis (LDA) and other manifold

learning methods have been applied to extract sensitive features for the patterns hidden in the high-dimensional data. KPCA is an extension of PCA by using the kernel trick. However, these methods transform high-dimensional data into low-dimension feature space based on the second order statistics, such as variance, correlation and mean square error. They choose the features depending on the top eigenvalues and corresponding eigenvectors of certain matrix without revealing the nature of original feature sets.

Kernel entropy component analysis (KECA), firstly proposed and employed in pattern recognition by Jenssen [8], is a novel nonlinear feature extraction method. It is developed based on information theory aimed at preserving the maximum Renyi entropy of the input data with minimum extracted features. However, KECA is an unsupervised method without considering the class information of the patterns, which may discard some useful information. LDA is a typical supervised method, which extract features by maximizing the ratio of the between-class distance and the within-class distance. To take full use of the labelled information, we proposed a supervised method combined KECA and LDA for fault diagnosis of rolling bearings inspired by Yang [9] and Lei's works [10]. Moreover, due to the information entropy and scatter information, it can reveal structure related to Renyi entropy of the input space data set and introduce discriminant classification information of different classes. Thus, it can improve recognition performance.

2. THEORETICAL BACKGROUND

The projection direction of KECA depends on the Renyi entropy, which extract features by maintaining the minimum loss of Renyi entropy of the input data after data transform. Assuming that $p(\mathbf{x})$ is the probability density function of a given sample $\mathbf{X}=\mathbf{x}_1, \dots, \mathbf{x}_N$, then its Renyi entropy of the order α is expressed as

$H_\alpha(\mathbf{X}) = \frac{1}{1-\alpha} \lg\left(\int p^\alpha(\mathbf{x}) d\mathbf{x}\right)$ [11], where $\alpha \geq 1$. In KECA, Renyi quadratic entropy, i.e., $\alpha = 2$ is employed for the entropy can be elegantly estimated by Parzen window density estimator [12], then Renyi quadratic entropy of the data set can be expressed by:

$$H(\mathbf{X}) = -\lg\left(\int p^2(\mathbf{x}) d\mathbf{x}\right) \quad (1)$$

Because the logarithmic is monotonous, we can only focus on the integral function $V(p) = \int p^2(\mathbf{x}) d\mathbf{x}$ [8, 13, 14]. To estimate $V(p)$, and hence $H(p)$, a Parzen window density estimator $\bar{p}(\mathbf{x}) = \frac{1}{N} \sum_{\mathbf{x}_i \in D} \mathbf{K}_\sigma(\mathbf{x}, \mathbf{x}_i)$ is introduced [8,12], where $\mathbf{K}_\sigma(\mathbf{x}, \mathbf{x}_i)$ is the estimator or kernel function centered at \mathbf{x}_i and σ is the kernel size. Substituted $\mathbf{K}_\sigma(\mathbf{x}, \mathbf{x}_i)$ and $\bar{p}(\mathbf{x})$ into $V(p)$, then $V(p)$ can be rewritten as follows:

$$\begin{aligned} \bar{V}(p) &= \int p^2(\mathbf{x}) d\mathbf{x} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int \mathbf{K}_\sigma(\mathbf{x}, \mathbf{x}_i) \mathbf{K}_\sigma(\mathbf{x}, \mathbf{x}_j) d\mathbf{x} \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{K}_{\sqrt{2}\sigma}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{N^2} \mathbf{K} \mathbf{1} \mathbf{1}^T \end{aligned} \quad (2)$$

here, \mathbf{K} is a $N \times N$ kernel matrix, the element (i, j) of \mathbf{K} is $\mathbf{K}_\sigma(\mathbf{x}_i, \mathbf{x}_j)$, and $\mathbf{1}$ is a $N \times 1$ vector containing ones. Therefore, the Renyi entropy can be estimated by the eigenvalues and eigenvectors of the constructed kernel matrix which can be eign-decomposed as $\mathbf{K} = \mathbf{E} \mathbf{D} \mathbf{E}^T$, where $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ and $\mathbf{E} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$. Here λ_i and \mathbf{a}_i are the i -th eigenvalues and corresponding eigenvectors, respectively. Then

$$\bar{V}(p) = \frac{1}{N^2} \mathbf{1} \mathbf{K} \mathbf{1}^T = \frac{1}{N^2} \mathbf{1} \mathbf{E} \mathbf{D} \mathbf{E}^T \mathbf{1}^T = \frac{1}{N^2} \sum_{i=1}^N \left(\sqrt{\lambda_i} \mathbf{a}_i^T \mathbf{1} \right)^2 \quad (3)$$

This expression is the so-called entropy values, and each term in Eq. (3) contributes to the entropy estimation. The eigenvectors and corresponding eigenvalues are ranked in decreasing order of the entropies.

KECA selects certain eigenvalues and corresponding eigenvectors according to the d largest entropies [8], different from PCA and KPCA that select largest eigenvalues. Therefore, the resulting KECA expression can be expressed as $\mathbf{Y} = \mathbf{D}_d^{-\frac{1}{2}} \mathbf{E}_d^T$, where \mathbf{D}_d and \mathbf{E}_d store the top d eigenvalues and corresponding eigenvectors.

The input data is transformed into the KECA space, which removes some redundant features and retains the distribution information of data. Next, the LDA algorithm is implemented in KECA space in search of optimal features that maximizes classification accuracy. The between-class scatter matrix \mathbf{S}_b^{keca} and the within-class scatter matrix \mathbf{S}_w^{keca} are expressed as:

$$\mathbf{S}_b^{keca} = \sum_{i=1}^c n_i (\mathbf{m}_i^{keca} - \mathbf{m}^{keca})(\mathbf{m}_i^{keca} - \mathbf{m}^{keca})^T \quad (4)$$

$$\mathbf{S}_w^{keca} = \sum_{i=1}^c \sum_{j=1}^{n_i} (\mathbf{x}_{ij}^{keca} - \mathbf{m}_i^{keca})(\mathbf{x}_{ij}^{keca} - \mathbf{m}_i^{keca})^T \quad (5)$$

where c is the number of classes, n_i is the number of training samples of i th-class, \mathbf{m}_i^{keca} is the mean of the samples belong to i th-class, and \mathbf{m}^{keca} is the overall mean of the samples.

Maximize the ratio of \mathbf{S}_b^{keca} and \mathbf{S}_w^{keca} , we can obtain the optimal projection direction \mathbf{W}^{keca} as follows:

$$\max J(\mathbf{W})^{keca} = \frac{(\mathbf{W}^{keca})^T \mathbf{S}_b^{keca} \mathbf{W}^{keca}}{(\mathbf{W}^{keca})^T \mathbf{S}_w^{keca} \mathbf{W}^{keca}} \quad (6)$$

where \mathbf{W}^{keca} contains a set of generalized eigenvectors of the eigen-equation $\lambda^{keca} \mathbf{S}_w^{keca} \mathbf{W}^{keca} = \mathbf{S}_b^{keca} \mathbf{W}^{keca}$ corresponding to d largest eigenvalues. Based on the above analysis, the transformed data $\mathbf{Z} = (\mathbf{W}^{keca})^T \mathbf{D}_d^{-\frac{1}{2}} \mathbf{E}_d^T$ processed by KECA-LDA can be obtained. The process of KECA-LDA can be summarized as follows:

Step1: Transform the input data into d -dimensional space by using KECA, the given sample \mathbf{X} is transformed to KECA-based feature vector $\mathbf{Y} = \mathbf{D}_d^{-\frac{1}{2}} \mathbf{E}_d^T$.

Step2: Construct the between-class scatter matrix \mathbf{S}_b^{keca} and the within-class scatter matrix \mathbf{S}_w^{keca} in the KECA transformed space.

Step3: Calculate the generalized eigenvectors $\mathbf{W}^{keca} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d)$ of eigen-equation $\lambda^{keca} \mathbf{S}_w^{keca} \mathbf{W}^{keca} = \mathbf{S}_b^{keca} \mathbf{W}^{keca}$ corresponding to d largest positive eigenvalues, and the transformed feature is $\mathbf{Z} = (\mathbf{W}^{keca})^T \mathbf{Y} = (\mathbf{W}^{keca})^T \mathbf{D}_d^{-\frac{1}{2}} \mathbf{E}_d^T$.

Step4: The extracted features are fed into support vector machine (SVM) classifier for further classification.

3. EXPERIMENT AND RESULTS

3.1. Data description

The raw vibration signals from the bearing data center [15] are used for evaluating the effectiveness of the proposed method for fault diagnosis of rolling bearings. The data has been validated in many studies and becomes a standard reference in the bearing diagnostics filed [16]. The test setup consists of four parts, i.e.,

a motor, a dynamometer, a torque transducer and control electronics, as shown in figure 1. The vibration signals were collected by the accelerometer amounted on the motor's shell. The sampling frequency is 12 kHz, the motor speed is 1797rpm under 0 horse power/hp. Single point faults of size 0.007, 0.014, 0.021 and 0.028 in. were seeded on the different parts of the drive-end bearings (SKF deep-groove ball bearings: 6205-2RS JEM) of the motor by using electro-discharge machining (EDM).

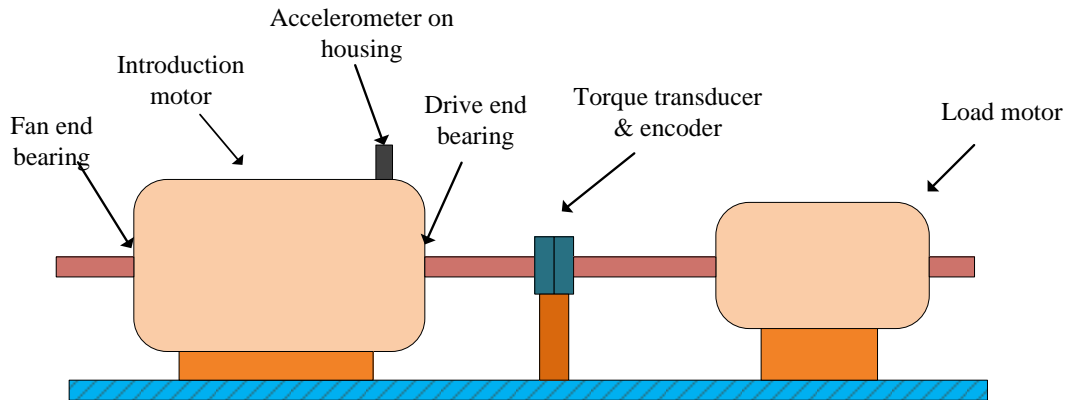


Figure 1. Test rig.

Four operating conditions (normal, rolling elements fault, inner and outer races faults) were introduced in the experiments. The last three fault types involved three different fault sizes of 0.007, 0.014 and 0.021 in. in this study. Thus, a ten-class classification problem is investigated, and the samples are split into 100 data per operating condition, 50 samples are randomly selected for training, and the remaining samples are used for testing, the detail description of the datasets are shown in table 1.

Table 1: Description of data sets for fault diagnosis.

| Number of training/testing samples | Fault type | Fault size of training/testing samples (inches) | Abbreviation | Label of class |
|------------------------------------|------------|-------------------------------------------------|--------------|----------------|
| 50/50 | Normal | 0/0 | normal | 1 |
| 50/50 | Inner race | 0.007/0.007 | I-0.007 | 2 |
| 50/50 | Ball | 0.007/0.007 | B-0.007 | 3 |
| 50/50 | Outer race | 0.007/0.007 | O-0.007 | 4 |
| 50/50 | Inner race | 0.014/0.014 | I-0.007 | 5 |
| 50/50 | Ball | 0.014/0.014 | B-0.007 | 6 |
| 50/50 | Outer race | 0.014/0.014 | O-0.007 | 7 |
| 50/50 | Inner race | 0.021/0.021 | I-0.007 | 8 |
| 50/50 | Ball | 0.021/0.021 | B-0.007 | 9 |
| 50/50 | Outer race | 0.021/0.021 | O-0.007 | 10 |

3.2. Results

Wavelet packet decomposition was performed to extract the fault features including the relative energy in a wavelet packet node and the entropy in a wavelet packet node [17]. Thus, the high-dimensional feature set containing 32 features was constructed. The high-dimensional feature set including the wavelet packet entropies may contain redundant information which increases the computation time and weakens the

recognition performance. To improve the accuracy, the KECA-LDA method was performed to feature reduction for removing the superfluous information. KPCA, KECA and KPCA-LDA were conducted for qualitative comparison to highlight the effectiveness of the proposed method. The target dimensionality for every method is set to a certain number so that the cumulative variance contribution rate is more than 95%. The results of the first three principal components were plotted after feature reduction with these methods, in order to display more intuitively, as depicted in figure 2. As can be observed, KPCA, KECA and KPCA-LDA couldn't separate the fault absolutely, where a small part of some classes are overlapped, especially for the different size of ball faults. By contrast, KECA-LDA has little misjudgment samples and obtains a fairly good and clear separation of the clustering results for those different operating conditions. The same clusters of KECA-LDA transformed samples are more compact and the distance between different classes is relatively far, which is conducive to the classification. It proves that KECA-LDA has better clustering performance than KPCA-LDA, because KECA-LDA combines the information entropy and discriminatory information. KECA-LDA is suitable as a feature extraction step prior to classification, and it functions well for classification to solve fault pattern recognition problem.

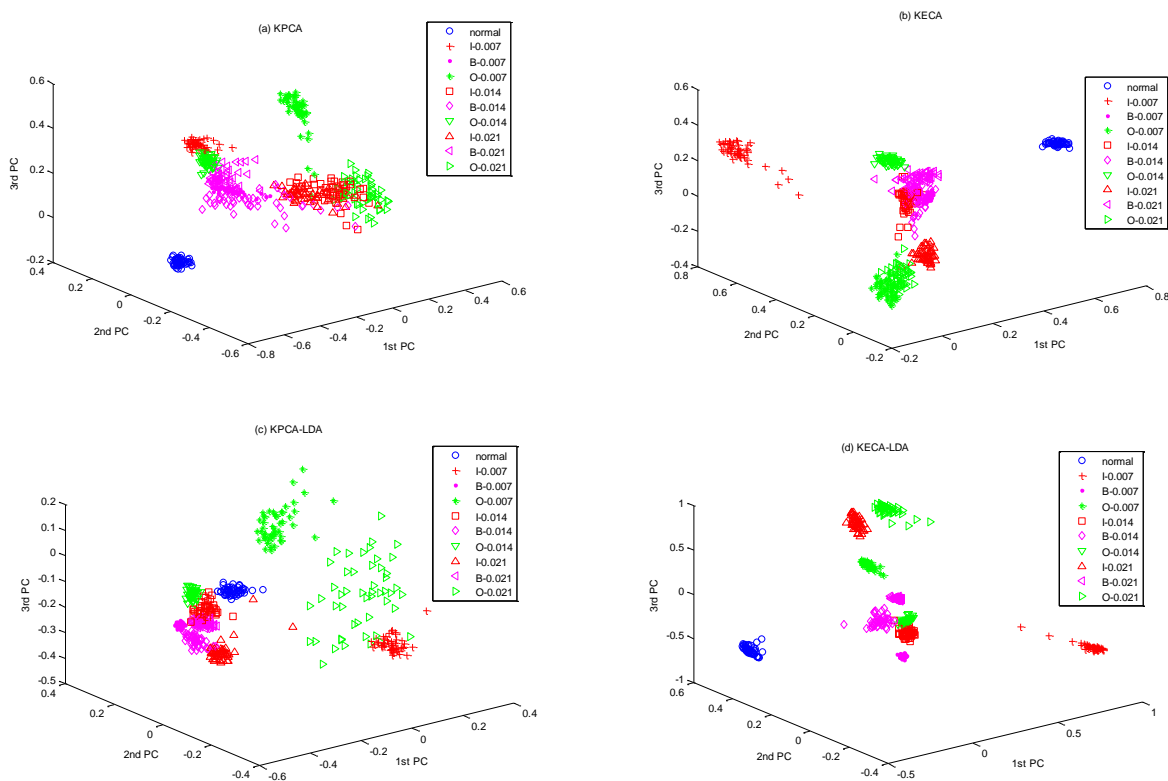


Figure 2. The 3-D subplots of KPCA-LDA and KECA-LDA for feature reduction: (a) KPCA, (b) KECA, (c) KPCA-LDA. (d) KECA-LDA.

Within the fault diagnosis of rolling bearings related to pattern recognition, in conjunction with feature reduction methods that find low-dimensional representative features for samples, classifiers need to be carried out to identify those different bearing faults. In this work, Support vector machine (SVM) is adopted for its well-developed statistical learning theory. The parameters of RBF kernel was optimized by the cross-validation method. The quantitative evaluation procedure has been repeated for 10 times in each experiment, and 50 training data per class is selected randomly each time. In order to highlight the effectiveness of the KECA-LDA, the classification accuracy of the proposed method was compared with the results of original features, PCA, KPCA, KECA, LDA and KPCA-LDA. The testing average results are shown in table 2, and the accuracies of SVM, PCA-SVM, KPCA-SVM, KECA-SVM, LDA-SVM, KPCA-LDA-SVM and KECA-LDA-SVM are 87.2%, 90.2%, 92.6%, 97.2%, 98.2%, 95.4% and 99.8%, respectively. The results

demonstrate that satisfactory overall classification results have been achieved by means of the dimension reduction. It can be seen that both KECA and KECA-LDA are effective on feature extraction and can improve the classification accuracy, while KECA-LDA has better classification performance than KPCA-LDA and KECA.

Table 2: The classification accuracies (%) of different methods with SVM classifier.

| Methods | original (%) | PCA (%) | KPCA (%) | KECA (%) | LDA (%) | KPCA-LDA (%) | KECA-LDA (%) |
|----------|--------------|---------|----------|----------|---------|--------------|--------------|
| Accuracy | 87.2 | 90.2 | 92.6 | 97.2 | 96.2 | 95.4 | 99.8 |

4. CONCLUSIONS

In this study, a new hybrid feature reduction method which combines KECA with LDA was proposed for fault diagnosis of rolling bearings. Based on information entropy and scatter information, KECA-LDA is effective for feature extraction and dimension reduction. A high-dimension feature set is first constructed, which contains the wavelet packet entropies derived from the raw vibration signals. Then KECA-LDA is used to dimensional reduction, which can extract the representative features and remove the superfluous information, meanwhile, maintain the intrinsic geometric structures. Finally, the extracted features of the feature space derived from the original vibration signals are fed into the SVM classifier to recognize different operating conditions of bearings. To verify the effectiveness of the proposed method, PCA, KPCA, KECA, LDA and KPCA-LDA are also conducted on the bearing set for comparison. Apparently, KECA-LDA has outperformed other methods in terms of achieving higher testing accuracies. The results demonstrate the feasibility and effectiveness of the proposed method for fault diagnosis of rolling bearings.

ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China (Grant No.s: 51575202, 51675204). The authors also thank the free download of the rolling element bearing fault datasets provided by Case Western Reserve University, Rockwell Science Office of Naval Research and CVX.

REFERENCES

- [1] Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32(8), 469-480.
- [2] Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96, 289-306.
- [3] Nandi, S., Toliyat, H. A., & Li, X. (2005). Condition monitoring and fault diagnosis of electrical motors-a review. *IEEE Transactions on Energy Conversion*, 20(4), 719-729.
- [4] Tang, B., Song, T., Li, F., & Deng, L. (2014). Fault diagnosis for a wind turbine transmission system based on manifold learning and shannon wavelet support vector machine. *Renewable Energy*, 62(3), 1-9.
- [5] Lei, Y., He, Z., Zi, Y., & Hu, Q. (2007). Fault diagnosis of rotating machinery based on multiple anfis combination with gas. *Mechanical Systems & Signal Processing*, 21(5), 2280-2294.
- [6] Yu, J. (2016). Machinery fault diagnosis using joint global and local/nonlocal discriminant analysis with selective ensemble learning. *Journal of Sound & Vibration*, 382, 340-356.
- [7] Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(10), 1299-1319.
- [8] Jenssen, R. (2009). Kernel entropy component analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 32(5), 847-860.
- [9] Liu, N., & Wang, H. (2012). Weighted principal component extraction with genetic algorithms. *Applied Soft*

- Computing Journal, 12(2), 961-974.
- [10] Gao, L., Qi, L., Chen, E., & Guan, L. (2014). A fisher discriminant framework based on Kernel Entropy Component Analysis for feature extraction and emotion recognition. IEEE International Conference on Multimedia and Expo Workshops (pp.1-6). IEEE.
 - [11] Rényi, A. (1961, June). On measures of entropy and information. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability (Vol. 1, pp. 547-561).
 - [12] Jenssen, R. (2009). Information Theoretic Learning and Kernel Methods. Information Theory and Statistical Learning. US: Springer. 209-230.
 - [13] Jiang, Q., & Shi, J. (2014). Sparse kernel entropy component analysis for dimensionality reduction of neuroimaging data. Engineering in Medicine & Biology Society (Vol.2014, pp.3366). Conf Proc IEEE Eng Med Biol Soc.
 - [14] Gao, L., Qi, L., Chen, E., & Guan, L. (2014). A fisher discriminant framework based on Kernel Entropy Component Analysis for feature extraction and emotion recognition. IEEE International Conference on Multimedia and Expo Workshops (pp.1-6). IEEE.
 - [15] Case Western Reserve University Bearings Data Center Website, <http://csegrouops.case.edu/bearingdatacenter/home>
 - [16] Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: a benchmark study. Mechanical Systems & Signal Processing, s 64–65, 100-131.
 - [17] Feng, Y., & Schlindwein, F. S. (2009). Normalized wavelet packets quantifiers for condition monitoring. Mechanical Systems & Signal Processing, 23(3), 712-723.

Drill Bit Wear Monitoring Based on Vibration Signal Analysis

Hamed Rafezi, Ferri Hassani

Department of Mining and Materials Engineering, McGill University, Montreal, Canada

ABSTRACT

This paper presents a novel approach for tricone bit wear monitoring based on vibration signal analysis in drilling for mining applications. An extensive data acquisition from full scale in-situ tests in participating mine sites in Canada was accomplished. A drill rig was equipped with data acquisition units and several accelerometers on different spots of the machine. Based on vibration signal frequency spectrum analysis, the characteristics of the pattern generated by worn bit are achieved. Accordingly, the frequency bands of vibration signal that are sensitive to bit wear are introduced. Furthermore, Wavelet Packet Decomposition (WPD) is applied on the vibration signal to present a time-frequency representation. At the next step, a best-tree for the worn bit signal representation is achieved. The final results of this research will make the mining companies able to detect the bit wear and also to predict the bit failure in order to avoid catastrophic failure and prevent production losses.

Keywords: Mining, Drilling, Tricone, Wear, Vibration

Corresponding author: Hamed Rafezi (email: hamed.rafezi@mail.mcgill.ca)

1. INTRODUCTION AND BACKGROUND

As mining industry is moving toward automation and increasing the efficiency and precision in production, a successful drilling condition monitoring is a vital step forward. Drilling and blasting are two preliminary tasks in large surface mining operations and constitute more than 15% of the total costs [1]. Tricone bits are preferred in most rotary drilling applications for blasthole drilling. In order to maximize the efficiency and placement of explosive energy, vertical or inclined blast holes are created using large hydraulically or electrically-controlled drill rigs. Tricone or fixed-type bits are implemented in rotary drill units. The fixed-type bits are limited to softer materials and penetrate in the rock by shearing it. Tricone bits (figure 1) are versatile and preferred in most rotary drilling applications, this study will focus on this type of bits.

The drill bit monitoring is important because the bit wear grade effects not only the drilling operation efficiency but also the rig maintenance costs and availability in long term. In addition, bit failure in the drill hole creates major delay to remove the detached cone(s) from the hole in order to avoid problems for the rock crusher equipment. Autonomous drilling would not be achievable without an understanding of when the drill bit is worn and requires a replacement.



Figure 1. A tricone rotary bit

For a better illustration of what is happening at the bit-rock interaction it is helpful to simplify the process down to the interaction of a single insert acting on a rock. This interaction could be representative of the interactions between the tricone rotary bit which includes many of inserts and the rock mass. Maurer introduced a model to show the basis of rock bit interaction in rotary drilling for a single insert [2,3].

Wear happens as a result of bit-rock interaction and major failure modes of tricone bits could be defined in three categories:

Gradual wear: The most basic type of wear which can happen on the teeth, cones as well as shirrtail structure

Teeth breakage: As a result of repeated impacts and fatigue in the harsh drilling environment, teeth breakage is a common issue in tricone drilling.

Bearing Failure: The effects of wear on the bearing elements as well as the edges of the cones and shirrtail structure will result in progressive wear in the rotary structure and eventually the bearing failure. This problem is known as catastrophic failure in drilling operation which will causes further costs and process interruption.

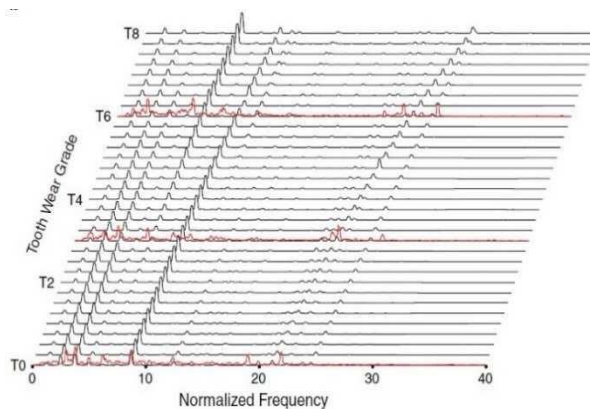


Figure 2. Vibration spectra for tri-cone rotary drill bit with different tooth wear grades [4]

The most well-known approach to evaluate bit wear is performance based method. The current state of bit wear is determined by monitoring penetration rate and torque. Assuming a constant rock type, and therefore a uniform rock strength, and constant operating conditions penetration rate will decrease over time as bit wear increases. Once the penetration rate has fallen below a pre-determined value the bit should be replaced [5].

The mentioned procedure is applicable for gradual tooth wear. For cone bearing failures the torque could be a Criterion. Any observation of a rapid rise in torque while under constant operating conditions is a sign of bearing failure. A bearing failure requires immediate bit replacement [5].

Unfortunately, the requirement of constant operating conditions and rock type is not the case in the open pit mining environment. Typically, several rock layers come across during the drilling of a single hole. Some layers may even be composed of entirely broken and non-homogenous rock material. At some particularly dynamic mine sites rock layers can be encountered in differing orders from one hole to the next within the same drill pattern. This

dynamic, and often unknown, in-situ geology makes it incredibly difficult to determine if penetration rate is decreasing or torque is spiking because of bit wear/failure or because a new rock layer is being encountered. This problem results in an essential limitation on performance based drill bit wear monitoring and is a motivation to explore new approaches and models for bit wear monitoring. Another work has been proposed which would attempt to correlate the bit performance data with drilling rock recognition data to estimate bit wear, however this has not yet managed to overcome problems to progress from the proposal stage [6].

Vibration signal is widely used in context of condition monitoring. The natural frequency of a blasthole drill's mast is observed to be in the 300-400 Hz range [7]. This is well above the vibration frequencies associated with tri-cone bit operation which is the part of interest for this study.

Drill string steel was found to have resonant frequencies of 75-80 Hz. While many of the higher frequency vibrations are caused by onboard rotating components such as motors and compressors [8]. For blasthole drilling application we can assume a rotation speed of between 40-120 rpm (revolutions per minute).

An initial study by Cooper showed that the vibration spectrum produced by a tri-cone rotary drill bit caused by the successive impacts of the bit teeth with the rock, could be associated with impacts coming from particular rows of teeth on a particular cone. Over time the vibration peaks were seen to accelerate as the drill bit teeth wore down [5]. The vibration frequency of a cone, in particular, accelerates with wear because the individual cone must move faster to cover the same distance as its size diminishes due to the wear. In addition to being an indicator of drill bit wear and bearing failure, vibration was also shown to have the potential to identify different rock types. Although Cooper's work seemed promising the conclusions did provide the warning that vibration data collection is practical in the lab but seems impractical and might be unfeasible in the field. However, this warning was directed at petroleum drilling which has extremely deep drill holes compared to the shallow holes seen in open pit mining. This has led to the belief that capturing vibration signals from the drill bit in an open pit mining environment would not be as challenging as in deep well drilling.

Other work explored the axial bit vibration spectra data to attempt to show a bit wear correlation. It was shown that tooth vibration peaks move to higher vibrations as wear progresses. Through continuous monitoring of axial bit vibration spectra detection of bit tooth wear can be accomplished from detecting spectral peak shifts, figure 2 [4]. In the figure T0 is the lowest tooth wear and T8 is the highest. The vibration peaks at around 20Hz accelerate from low tooth wear to high. These results were consistent with the previous vibration related work, providing more elaborate and an overall improvement. It must be noted that this testing was only accomplished in the laboratory setting and it is expected that in field drilling data will be much noisier and more confusing to interpret because of the effect of vibrations other than axial and therefore more difficult to interpret. Application of artificial neural network is also investigated [9].

The current research is developing a novel approach for tricone bit wear monitoring at based on vibration signals analysis in drilling for open pit mining applications. Present work is developed based on an extensive data acquisition from full scale in-situ tests in participating mine sites Canada. Several field works have been conducted and drill rigs have been equipped with accelerometers in various spots to collect the vibration data. The vibration signal during the whole life cycle of several bits at various working conditions have been analyzed in time and frequency domains to find the vibration pattern corresponding to the worn bits. Afterward, wavelet packet analysis is applied on the signals to achieve a time-frequency understanding of the signals behavior and to present a best tree to represent worn bit vibration signal.

2. EXPERIMENTS

The drilling tests were conducted in an iron ore mine and a copper mine in Canada. A BUCYRUS 49HR drill machine (figure 3) was equipped for the tests. Two triaxial accelerometers were mounted on the base and middle of the drill mast. Two data acquisition units (figure 4) were installed on the drill to digitize and record the vibration signals as well as following drilling signals:

- Hoist Motor Current
- Rotary Motor Current
- Head Encoder
- Bailing Air Pressure
- Hoist Voltage Request
- Rotary Voltage Request

The last three signals are drilling set points to be controlled by the operator. Bit diameters included 10 5/8" for trimming and buffering (T and B) holes and 12 1/4" bit for production patterns. Both air cooled and sealed bearing types were investigated.

Working on different benches and several patterns provides the drilling data in a wide variety of drilling parameters and geology conditions, and the aim of this work was to identify the signal features that are effected only by the bit wear and have a meaningful trend as the bit condition changes from a brand new bit totally worn out bit.



Figure 3. Test drill rig



Figure 4. Installed data acquisition units

3. SIGNAL ANALYSIS AND RESULTS

It was aimed to find a correlation between bit wear progress and drilling signals behavior and to detect signals statistical features and frequency components which are sensitive to bit wear and have a meaningful trend as the bit becomes worn. To achieve this purpose, for all the signals, the trends of frequency spectrum of vibration signals were analysed using Fast Fourier Transform (FFT). By comparison frequency trend of signals in different working conditions during the whole life cycle of the bits, it was found that specific frequency bands are sensitive to bit wear and have an incremental trend as the bit becomes worn.

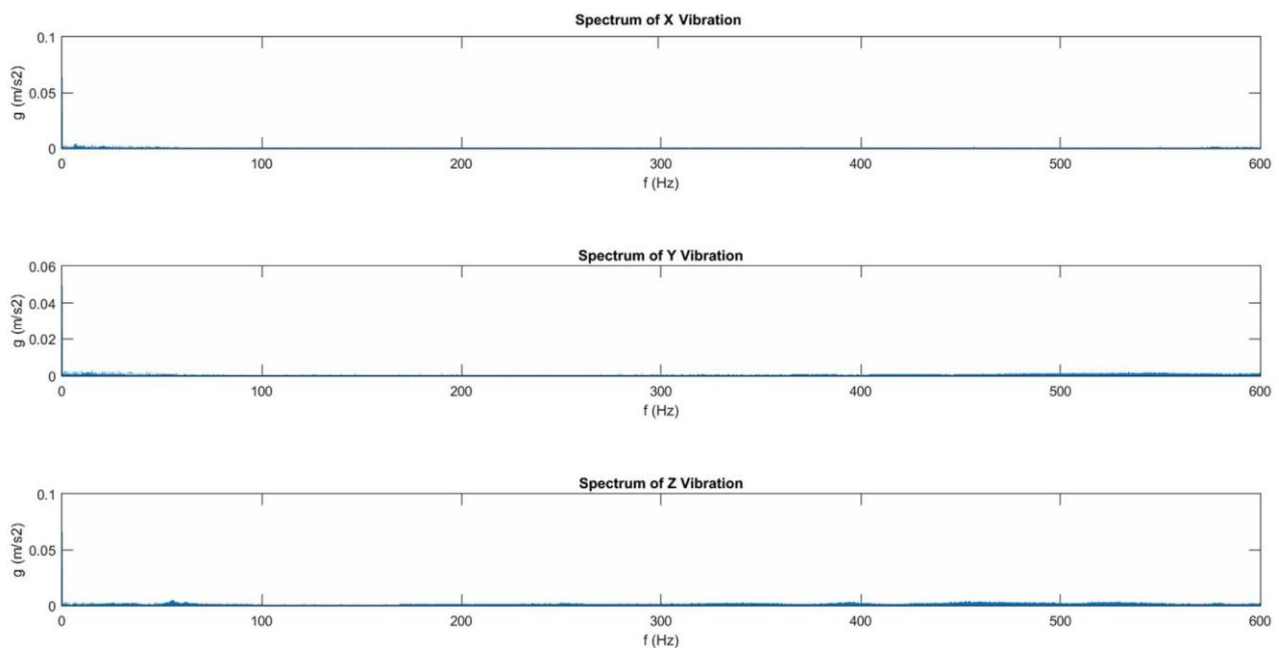


Figure 5- Horizontal (X,Y) and longitudinal (Z) vibration signal frequency spectrum of new bit

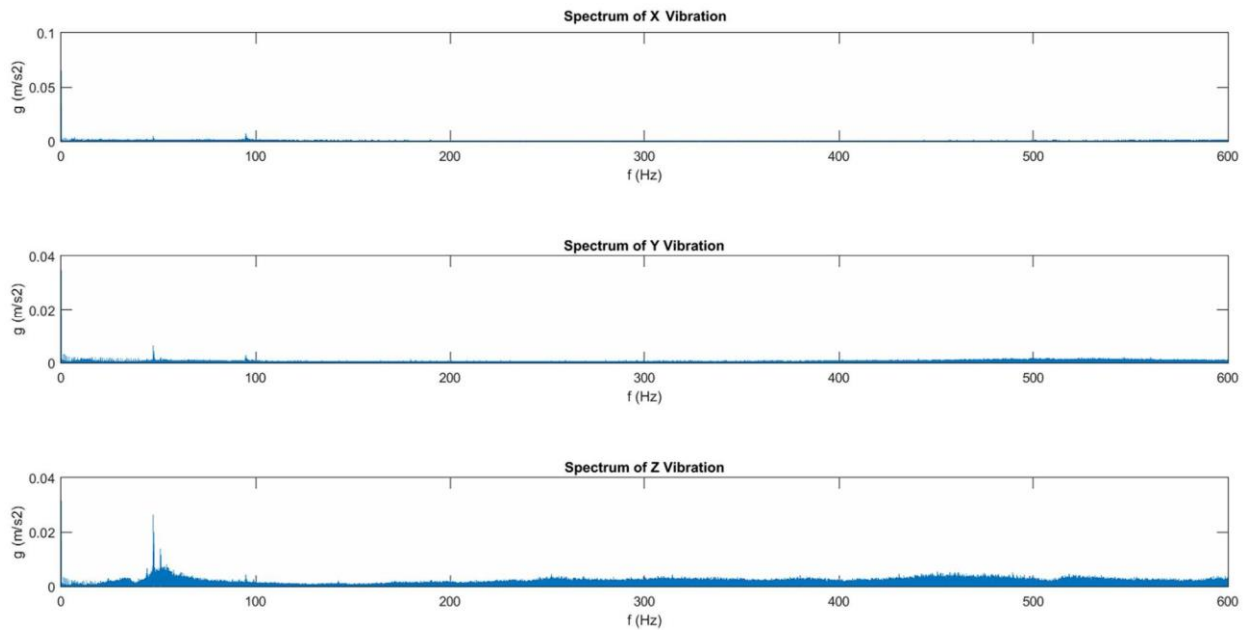


Figure 6- Horizontal (X,Y) and longitudinal (Z) vibration signal frequency spectrum of worn bit

Figures 5 and 6 illustrate the trend of frequency spectrum pattern as the bit transits from brand new to worn bearings condition. As shown, frequency components around 45 Hz and its harmonic i.e. 90 Hz are excited by bit with worn bearings and reach to their maximum value just before bit failure. This phenomenon is observable in both longitudinal and lateral vibration frequency spectrums. In the next step, a vibration signal has been analyzed in time-frequency domain using Wavelet Packet Decomposition (WPD). WPD works a series of low-pass high-pass filter and decompose the signal into lower and higher frequency components in every level of decomposition. In present work, Daubechies 5 is used as mother function and five levels of decomposition are investigated. Next step of this research includes designing a Artificial Neural Network (ANN) to classify the bit condition based on the vibration signal trend. For this purpose, due to huge volume of drill data a minimal representation of worn bit signal decomposition is achieved based on Shannon entropy criteria using MATLAB software.

Figure 7 shows the complete decomposition tree of the signal in five levels. After analyzing the worn bit vibration signals it was found that Packet (1,1) at first level, Packet (3,2) at third level, Packets (4,2), (4,3), (4,6), (4,7) at forth level and Packets (5,0), (5,1), (5,2), (5,3) at fifth level of decomposition represent the signal the best (figure 8). Having this minimal representation of the signal would yield in decrease required computational time in decision making at next stages of the project.

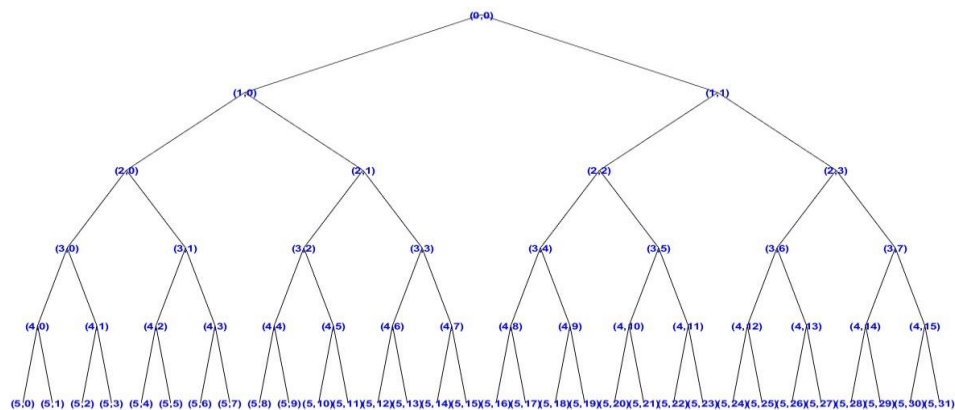


Figure 7- Initial WPD tree at five level of decomposition

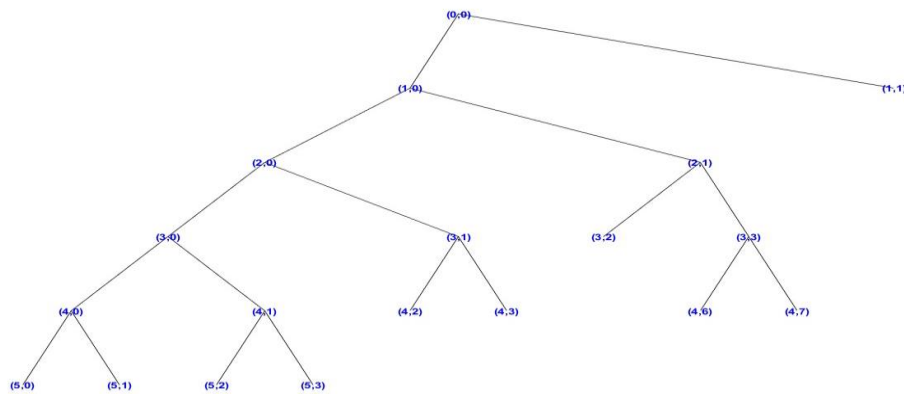


Figure 8- Best tree at five level of decomposition

4. CONCLUSIONS

This paper presents a part of a drill wear monitoring project at McGill University. The project consists of an extensive drilling data acquisition and data analysis based on full scale field test in participating mine sites in Canada. The real world data basis of the project makes it practical and reliable for industrial applications. The frequency spectrum pattern and specific frequency bands of vibration signal sensitive to bit wear are achieved and introduced. In addition, based on WPD analysis the best tree at five level of decomposition is introduced for further analysis in classification stages.

REFERENCES

- [1] Gokhale B.V., Rotary drilling and blasting in large surface mines (2011), ISBN 978-0415878784.
- [2] Maurer, W., Impact Crater Formation in Sandstone and Granite. Master's Thesis, Colorado School of Mines, (1959).
- [3] Maurer, W., The Perfect Cleaning Theory of Rotary Drilling. Jour. Pet. Tech. (1962) 1270-1274.
- [4] Naganawa, S., Feasibility study on roller-cone bit wear detection from axial bit vibration. Journal of Petroleum Science and Engineering 82-83 (2012) 140-150.
- [5] Cooper, G.A. et al., The Interpretation of Tricone Drill Bit Vibrations for Bit Wear and Rock Type. RETC Proceedings, Volume 1 (1987) 202-218.
- [6] Cooper, G.A., A proposal for the real-time measurement of drill bit tooth wear. Paper presented at Geothermal Resources Council Annual Meeting, Reno, Nevada, Sept. 22-25, (2002).
- [7] Branscombe, E.A. Investigation of Vibrated Related Signals for Monitoring of Large Open-Pit Rotary Electric Blasthole Drills. Master's Thesis, Queen's University, (2010).
- [8] Aboujaoude, C.E., Modeling, simulation and control of rotary blasthole drills. Master's Thesis, McGill University (1991).
- [9] Sidney Ly, J. J. Choi, Drill condition monitoring using ART-1, 0-7803-1901-X/94-IEEE (1994).

Detection and Diagnosis of Reciprocating Compressor Faults Based on Modulation Signal Bispectrum Analysis of Vibrations

Usama Haba, Khaldoon Brethee, Osama Hassin, Fengshou Gu and Andrew D. Ball
Centre of Efficiency and Performance Engineering, University of Huddersfield, Queensgate, Huddersfield
HD1 3DH, UK
E-mail: usama.haba@hud.ac.uk

ABSTRACT

Reciprocating compressors (RC) are one of the most common machines used in various industrial systems such as petroleum refineries, petrochemicals, chemical and refrigeration plants, etc. They usually provide a wide range of operating conditions and some of its components work under a harsh environment such as high pressure and high temperature which causes failures to RC components. Fault detection and diagnosis in RC at early stages not only provide useful information about the nature of the fault but also avoids great damage to mechanical systems and allows the RC to run at optimal efficiency.

This paper studies the vibration characteristics of different fault conditions when a RC operates under a wide range discharge pressures. Specifically, modulation signal bispectrum (MSB) is used to analyses the periodic couplings between structure resonances and operating speeds, which allows a more detailed representation of the nonlinear vibration responses due to mechanical impacts of valve plates upon their seats, fluctuation of airflow and nonlinear transfer paths and thereby obtain a stable and accurate features to indicate the root of the faults.

Experimental studies carried out for different common RC faults, such as valve leakage, intercooler leakage shows that average MSB magnitudes values can be used to differentiate the healthy condition from other faulty cases. The results show that MSB produces a more reliable and better performance in extracting fault features for more accurate detection and diagnosis different RC faults compared with spectrum analysis.

Keywords: Reciprocating compressor, fault detection and diagnosis, vibration signal, modulation signal bispectrum.

Corresponding author Mr. Usama Haba (Usama.ahab@hud.ac.uk)

1. INTRODUCTION

Reciprocating compressors as a useful power source are one of the most common machines used in industry because of their powerful design, flexible in use and high thermo-efficiency. These machines play an important part in different systems such as petroleum refineries, petrochemicals, chemical and refrigeration plants, etc., and offer a high value of reliability. Due to the complexity of the compressor structure and working condition, many of its components work in high-temperature and high-pressure environment [1] which endure reciprocating forces. Thus, these machines are vulnerable to various faults related to their functionalities and operational environments. Such faults not only degrade performance, consume additional energy, but also possibly result in an additional cost and even lead to severe consequences [2]. The valve is the most frangible part of reciprocating compressors, in which faults occur most frequently compared with other compressor parts. Thus cracked or broken RC valves as considered the biggest source of machines failures, are the most common reason for unscheduled shutdowns, can lead to secondary damage to other parts, reducing performance and serious machine breakdown [3-5]. Furthermore, deterioration of seals or intercooler joints also leads to intercooler leakage, this, in turn, will reduce compressor efficiency.

In practice, different faults may develop simultaneously causing damage to machines which increasing maintenance costs and reducing the plant efficiency. Therefore, efficient and effective condition monitoring

techniques are highly necessary and crucial to detect and diagnose machines faults at an early stage in terms of accident prevention, decision-making and reduce maintenance cost [6, 7]. Of many different monitoring techniques in developing, vibration analysis has been found to be the most popular and effective technique used in monitoring various faults in industrial machines includes reciprocating compressors such as valve leakage and intercooler leakage because of their visual feature, easy measurability and reliability [8]. To realise the early fault diagnostics, vibration mechanisms of a reciprocating compressor have been studied extensively.

Naturally, reciprocating compressors are noisy and vibrating, which commonly subjected to different failure modes such as fatigue, wear, and deformation. These failures can cause shaft misalignment, unbalances rotors which increase the vibration levels produced by the machine. Compressor-generated signals are normal symptoms of additional unbalanced, dynamic loads which may lead to machine breakdown [9]. These vibration signals are characterized by a series of periodic events such as piston slap, valve opening and closing impacts. This type of vibration signal is described as non-linear, non-stationary signal which statistical properties change periodically with time [8]. The impacts of compressor valve plates upon their seats normally generates transient vibrations. Additionally, fluctuation of airflow also produces vibration and sound. Such vibrations are so broad in their frequency range, so vibration spectra can be very complicated [3]. Since the vibration of the cylinder head is a summation of many excitations, including valves impacts during closing and opening operations, thus the excessive vibration spectrum is expected to be complex.

More attention is paid for compressor faults diagnosis, which has been studied for decades, and a large amount of research has been conducted and proposed. From all these approaches proposed in the literature, vibration analysis is the most beneficial method due to their reliability and high accuracy. Based on vibration signature analysis, different signal processing techniques such as Wigner–Ville distribution [10], wavelet transform [11, 12], neural networks [12, 13] and support vector machines [14, 15], have been used for fault detection and diagnosis of reciprocating compressors. Furthermore, valve leakage is the most frequently fault occurs in the reciprocated compressor. In which, various compressor valve leakages have been diagnosed using instantaneous angular speed [16, 17], relevant vector machines based on envelope spectrum analysis [2] and time-frequency representation [18] under different operating conditions.

Although the above-mentioned fault diagnosis methods have good performance in diagnosing compressor faults. However, most of them are based on statistical learning theory and data-driven from wideband data, which may include slightly unrelated information that leads to an unstable indication for incipient fault. This paper presents a new method for RC fault detection based on modulated signal bispectrum (MSB) analysis of surface vibration. MSB is an advanced signal processing technique used to capture the deterministic nonlinear feature of modulation components for a reliable diagnosis. The result are obtained from common power spectrum and MSB analysis applied to signals from a laboratory experimental setup operating under different compressor discharge pressure

2. THEORETICAL BACKGROUND

2.1. Compressor torque

Much more than with rotating machines, the vibration signals from reciprocating machines are a series of responses to impulsive events within the cycle of the machine, such as piston slap, valves opening and closing events, and so on [8]. Due to its reciprocating movements, RC always have some degree of torsional vibration during operation. While the crankshaft rotates, the piston goes through different strokes. The rotation of the crankshaft during compression stroke causes the piston to move upwards approaching the top dead centre (TDC) which increases the in-cylinder pressure. Whereas, when the piston starts to move downwards, the in-cylinder pressure starts to decrease until the piston reaches the bottom dead centre (BDC). The pressure difference across the piston normally generates tangential force (via the con-rod) that decreases the rotational speed of the crankshaft during the compression stroke leading to the crankshaft speed fluctuations. Predominantly, any change in the pressure will affect the rotational speed of the

crankshaft. Thus, faults occur in a cylinder or valve system (small or large) would affect the torsional vibration and leads to speed fluctuation [19].

Figure 1(a) shows compressor first and second stages cylinders pressure distribution synchronized with crankshaft angle and Figure 1(b) illustrates a typical two-stage piston-cylinders system used in reciprocating compressors. It consists of a piston with a length of connecting rod l crank and radius r . The compressor cycle starts at TDC where $\theta = 0^\circ$ and ends with $\theta = 360^\circ$ of the piston in the low-pressure cylinder. From the 90° V-shaped compressor construction, the high-pressure cylinder leads to the low-pressure cylinder by $\pi/2$. This phase difference is taken account into the equations applying to the high-pressure cylinder.

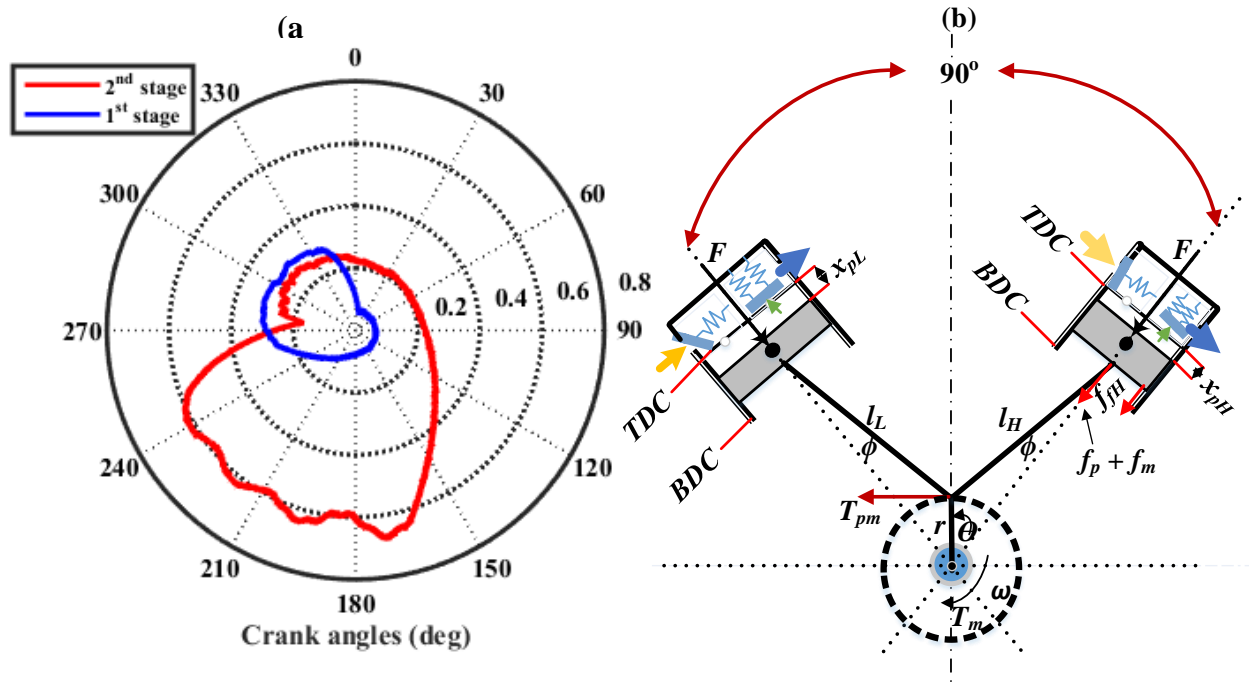


Figure 1: (a) Pressure distribution with crankshaft angle, (b) Simplified model of two stage air compressor

The piston is driven by the torque of the crankshaft in a reciprocating motion, via a connecting rod compressing the trapped air which exerts a force F directed against the motion of the piston (in both cylinders), which creates a torque M_t to rotate the crankshaft in a counterclockwise direction about Z. [3].

$$M_t = \left(\frac{F}{\cos \phi} \right) r \sin(\theta + \phi) \quad (1)$$

The effect of the resultant torque M_t is in both the low or high pressure compressor cylinders.

$$M_t = \begin{cases} T_{pmL} = (f_p + f_m)_L r \\ T_{pmH} = (f_p + f_m)_H r \end{cases} \quad (2)$$

$$T_{pmL,H}(t) = T_{pmL} + T_{pmH} \quad (3)$$

where $T_{pmL,H}$ is the resultant torque of the air pressure inside the cylinder and the unbalanced inertial force and the connecting rod of the low or high pressure cylinders. f_p is the tangential force produced by the air pressure in the cylinder, f_m is the tangential force produced by the inertia of the reciprocating mass, and r is the radius of the crank. The tangential force produced by the air pressure in a cylinder can be obtained from:

$$f_p = p_c s_c \left(\sin \theta + \cos \theta \frac{\left(\frac{r}{l}\right) \sin \theta}{\sqrt{1 - \left(\frac{r}{l}\right)^2 \sin^2 \theta}} \right) \quad (4)$$

where P_c is the pressure in cylinder, s_c is the cross-sectional area of the cylinder which can be calculated as following:

$$s_c = 0.25\pi d^2 \quad (5)$$

Where d is the bore diameters of the cylinder. The tangential force produced by the vertical inertial force for both cylinders becomes:

$$f_m = -m_{rec} \ddot{x}_p \left(\sin \theta + \cos \theta \frac{\left(\frac{r}{l}\right) \sin \theta}{\sqrt{1 - \left(\frac{r}{l}\right)^2 \sin^2 \theta}} \right) \quad (6)$$

The compressor inertial mass for both cylinders m_{rec} , which can be calculated as following

$$m_{rec} = m_p + 0.5m_{cr} \quad (7)$$

where m_p is the pistons mass, m_{cr} is the equivalent reciprocating mass of the connecting rod. x_p is the vertical piston acceleration and can be obtained based on the dynamic displacement motion of the piston as denoted in Figure 1(b):

$$x_p = r(1 - \cos(\theta)) + l \left(\sqrt{1 - \frac{r^2}{l^2} \sin^2(\theta)} \right) \quad (8)$$

The full torque description of the crank shaft can be derived according to Newton's second law [3]:

$$J \frac{d^2 \omega}{dt^2} = T_{em}(t) - T_{pmL,H}(t) - T_{fL,H}(t) \quad (9)$$

where ω is angular speed of the crankshaft, J is the equivalent inertial moment of the system, $T_{em}(t)$ is the driving motor torque and $T_{fL,H}(t)$ is the friction torque of the low pressure and high pressure cylinders. In which, the friction effect can be neglected and the equation yields:

$$d\omega_m = \frac{1}{J} [T_{em}(t) - T_{pmL,H}(t)] dt \quad (10)$$

2.2.Modulation signal bispectrum

The compressor normally works under different range of discharge pressure. This change in operating condition might cover the small changes caused by a fault. Thus separate different types of RC faults will be very difficult. Additionally, different faults may produce similar change patterns in both waveform and spectra which means that the second-order analysis is not sufficient to suppress random noise and distinguish these changes. Therefore, more advanced signal processing techniques have to be used to enhance the small changes for differentiating various types of compressor faults. The modulation signal of vibrations is formed by nonlinear coupling between vibration components such as, valve plates upon their seats, either in periodic or random forms. Therefore, it is expected that this bispectrum can give a more

accurate representation of the RC vibration signal for early detection of abnormal operations. For a discrete time current signal $x(n)$ its discrete Fourier Transform (DFT) $X(f)$ is defined as:

$$X(f) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi fn} \quad (11)$$

The well-known second order measure, the power spectrum (PS) of $x(n)$ is computed by the formula

$$P(f) = E[X(f)X^*(f)] \quad (12)$$

where $X^*(f)$ is the conjugate of $X(f)$ are the Fourier transform of the signal sequence $x(n)$ and $E[\cdot]$ is the statistical expectation. The power spectrum is a linear transform and it contains only amplitude information of individual component f . ignoring the effects of signal phases, which leads to inclusion of random noise. Extending these definitions to measures the third order gives rise to the bispectrum $B(f_1, f_2)$ which can be defined in the frequency domain as [6, 20]

$$B(f_x, f_c) = E[X(f_x)X(f_c)X^*(f_x + f_c)] \quad (13)$$

where f_x, f_c and $(f_x + f_c)$ are three individual frequency components achieved from Fourier series integral. Equation (3) includes only the presence of nonlinear effects from the harmonically related frequency components f_x, f_c , and $(f_x + f_c)$ and it fails to notice the possibility that the occurrence of $(f_x - f_c)$ might be due to the presence of quadratic phase coupling (QPC) between the two components f_x and f_c . For this reason, it is not accurately enough to describe the amplitude modulation (AM) signals such as motor current signals. To enhance the performance of the conventional bispectrum to characterize the motor current signals, a modified bispectrum method, named as modulated signal bispectrum (MSB) was proposed in [20-23] as following

$$B_{MS}(f_x, f_c) = E[X(f_c + f_x)X(f_c - f_x)X^*(f_c)X^*(f_x)] \quad (14)$$

where f_x is the modulator frequency; f_c is the carrier frequency. $(f_c + f_x)$ and $(f_c - f_x)$ are the higher and lower sideband frequencies respectively and now both considered in equation (4) simultaneously for measuring the nonlinear effects of modulation signals. If both $(f_c + f_x)$ and $(f_c - f_x)$ are due to non-linear effect between f_x and f_c then a bispectral peak will appear at bifrequency $B_{MS}(f_x, f_c)$. On the other hand, if these components for instance, various noises are not coupled but have random distribution, then their MSB magnitude will be close to zero. Thus, the aperiodic components of vibration signals and uncorrelated wideband noise can be suppressed effectively. Therefore, the discrete components relating modulation effects can be represented sparsely and characterised more effectively [24].

3. EXPERIMENTAL PROCEDURE

To evaluate the analysis in the previous section, an experimental study was employed based on a two-stage, single action V-shaped Broom Wade TS9 reciprocating. It can deliver compressed air up to 0.8MPa (120 psi) to a horizontal air tank. The compressor is driven by a three phase KX-C184, 2.5kw induction motor via a transmission belt as shown in figure 2. In the test, three working conditions were performed one by one: healthy condition (BL), discharge valve leakage on the second stage (DVL), leakage on the intercooler pipeline (ICL) connecting the 1st stage discharge valve with the 2nd stage suction valve.

During this experimental work, all the data was recorded using the Power 1401 Plus CED high-speed data acquisition system which converts the analog voltage into digital values, and it can record waveform data, event data and marker information at 400 kHz and resolution at 16 bit [25]. This system has 16 channels; for collecting data only seven channels were used. The maximum sampling frequency is 49 kHz, and each test recorded data at 3.6 second intervals. A speed encoder is mounted on the crankshaft that produces squared pulse output per revolution for measuring the IAS of the compressor drive shaft.

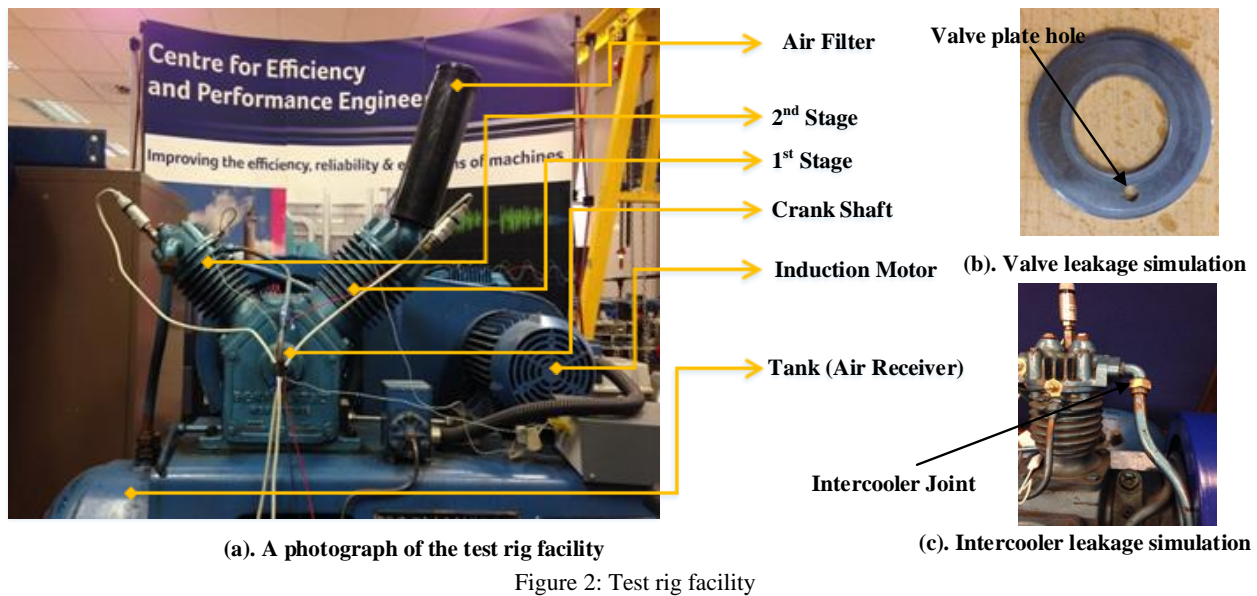


Figure 2: Test rig facility

4. RESULTS AND DISCUSSION

From both the 1st and 2nd stages vibration and cylinder pressure comparison in figure 3, it can be seen the simulated faults have caused observable changes on the compressor performance.

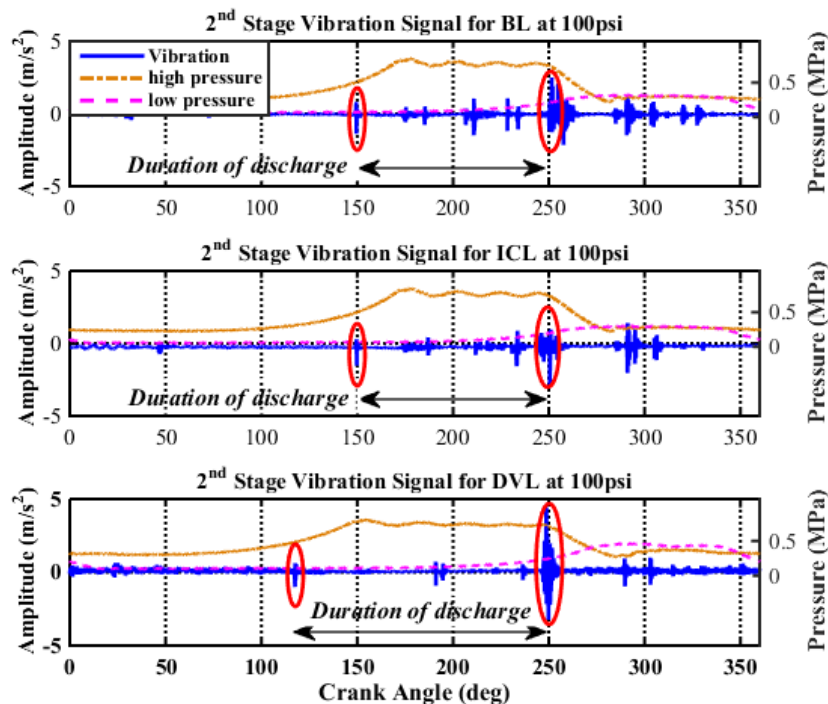


Figure 3: Measured values of low and high cylinder pressure & vibration signals for different cases at 100psi

Leakage through 2nd stage discharge valve allows a high pressure air to be pushed through the valve causing the 2nd stage cylinder pressure to build-up earlier than the healthy operation. Consequently, the discharge valve opens noticeably earlier compared to the healthy condition (as illustrated by red circle around the opening and closing of discharge pressure valve). These changes are due to the back flow through the discharge valve leakage and hence the discharge process occurs earlier. The larger the leakage size, the greater the leakage back into the cylinder and hence higher pressure in the suction process

4.1. Characteristics of vibration spectrum

The compressor vibration responses are characterized by a series of periodic events such as the piston reciprocating movement, mechanical impacts of valve plates upon their seats, fluctuation of airflow and nonlinear transfer paths. During the operation of the compressor, the mechanical movements of the valve (opening and closing impact) will normally exhibit metal-to-metal pattern which leads to transient vibrations. This pattern tends to generate shock pulses, which leads to different types of spectral features. i.e.; low-frequency harmonics describe the reciprocating movements and higher frequency harmonics for the valve impacts with frequencies in the range between 1kHz-10kHz [3]. Analyses these periodic couplings between structure resonances and operating speed will allow a more detailed representation of the non-linear vibration responses.

Figure 4(a) depicts the low-frequency spectral analysis of the measured vibration signals for different operational cases. When a leakage introduced into the second stage discharge valve, the spectrum of the impact noise shows significant differences in the amplitude of the vibration signal, the larger the leak the great difference with the healthy condition. Note that the vibration signal amplitude for the leaky valve substantially higher compared to the healthy operation. Additionally, the amplitude increases as the compressor discharge pressure increases. Figure 4(b) and (d) illustrate the first three harmonics amplitude change over compressor discharge pressure for healthy and simulated cases as symbol harmonic amplitude for the corresponding pressure condition for all cases. It can be seen very clearly that the amplitude for both healthy and faulty condition take similar shape and has the rising trend as the discharge pressure increases especially for the fundamental harmonic. Additionally, a significant difference in their amplitude (especially from 80psi) can be easily seen as the discharge valve leakage fault being the highest one, followed by intercooler leakage and the healthy condition being the lowest one. Figure 4(c) shows the 3rd harmonic amplitude, its shows a substantial difference between healthy and various fault cases. However, it is difficult to separate intercooler leakage from healthy condition. This means that the vibration signature from the low-frequency harmonics contains a sufficient information for compressor faults diagnosis and able to differentiate faulty cases from healthy operating with a clear separation

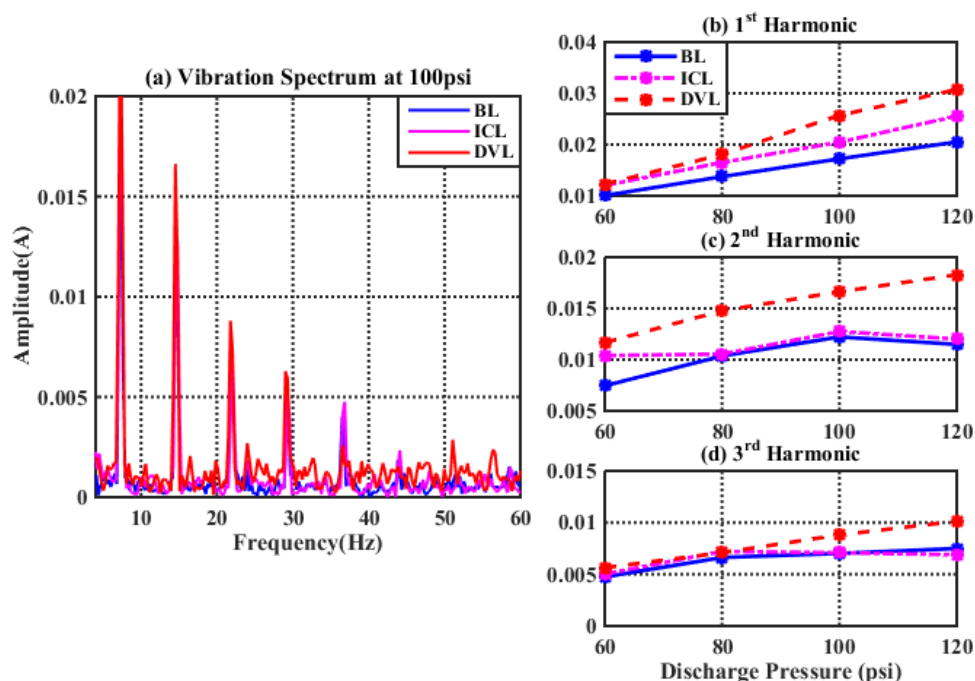


Figure 4: (a) 2nd stage vibration Spectrum, (b), (c) and (d) the First, second and third harmonics respectively for healthy and simulated cases.

4.2. Characteristics of MSB Magnitude

To avoid the influences of low-frequency vibration from other interfering vibration sources such as driving motor, MSB calculation was carried out in the high-frequency range for the carrier, $f_c = 3.0\text{kHz} \approx 4.5\text{kHz}$, whereas the modulator frequency $f_x = 0\text{Hz} \approx 30\text{Hz}$, in order to catch the shaft rotation frequency (7.4Hz) up to the fourth harmonics.

Figure 5 shows MSB magnitude results based on vibration signals for compressor performance under different discharge pressure with valve leakage and intercooler leakage. Each of subplots was obtained with a frequency resolution of 0.7480Hz and through 84 averages to ensure a stable result and the MSB magnitudes remain the same when further average was added.

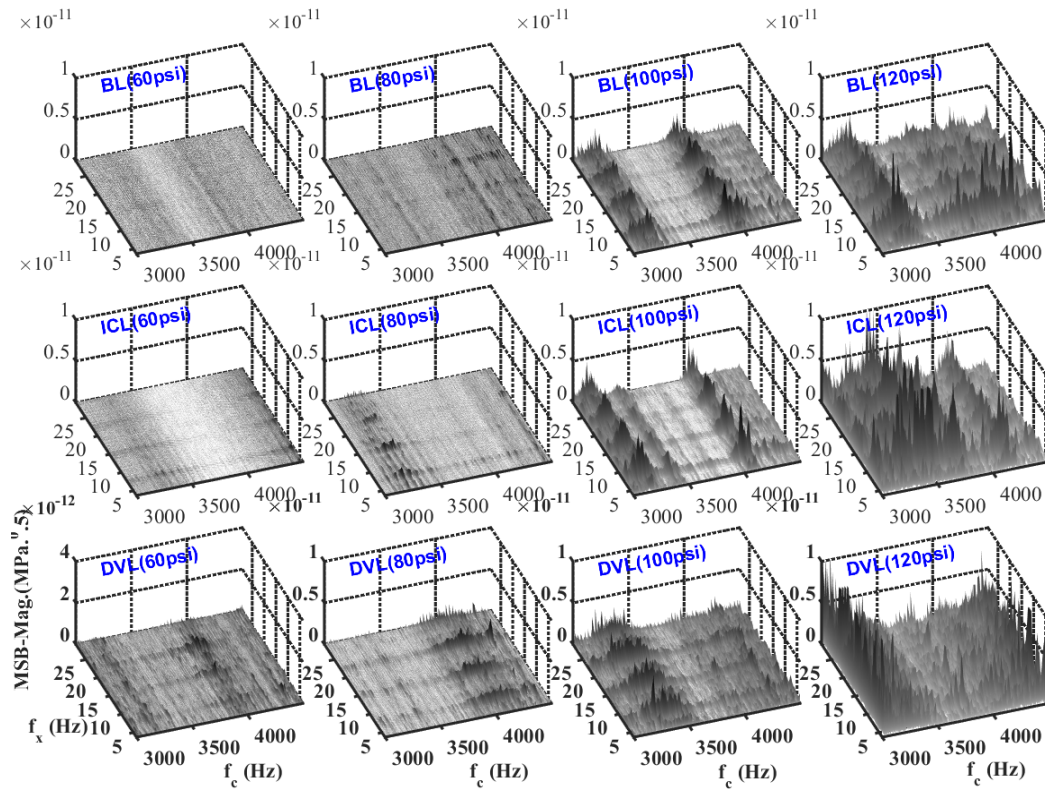


Figure 5: MSB Magnitude for different cases and different discharge pressure

Examine the figure representation it can be found that the MSB shows distinctive peak patterns across the carrier frequency due to the flutter of the valve plates and the fluctuation of airflow during the suction and discharge procedures particularly for the faulty cases. More importantly, the peak amplitude for simulated cases are more than (30%) higher than that of the healthy case especially at higher discharge pressure, which can be directly used as an important feature to detect the presence of these faults. Another observation is that, the MSB magnitudes are more continuous in shape across the modulator frequency and in different bands of the carrier frequency. As it appears in high discharge pressure conditions, it indicates the overwhelming effect of the valve plate collisions upon their seats due to highly pressured air passing through the valve system.

Experimental evaluation shows that the modulation signal bispectrum (MSB) magnitudes bring forward distinctive characteristics of vibration contents regarding to compressor conditions, providing useful information for diagnostics. Figure 6 illustrates the diagnostic performance of vibration signal for the Averaged MSB magnitudes for different faulty cases and under different discharge pressure. Based on the amplitude in the high frequency band as shown in figure 6 it is possible to differentiate the valve leakage fault from intercooler leakage and healthy condition as it happens at higher discharge pressure, valve plates flutters more, and thereby more collisions between plates and valve seats causing more and more vibration signatures.

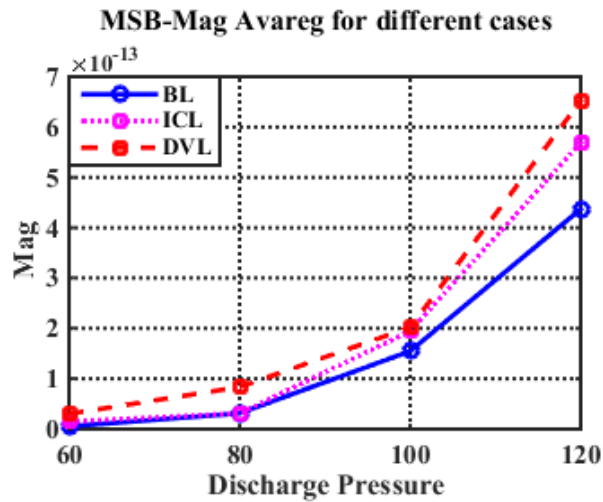


Figure 6: Averaged MSB magnitudes in the high frequency bands

5. CONCLUSION

Compressor valves work under a harsh condition such as high pressure and high temperature and considered as the main source of compressor failures. Based on the analytic studies and experimental verifications, it has been demonstrated that popular vibration monitoring method can detect valve leakage and intercooler leakage in a wide compressor operating conditions. Various signal processing techniques can be used with vibration signal, such as modulation signal bispectrum to extract useful information for reliable diagnosis. The MSB has the ability to characterise the nonlinear vibration effects of valve leakage and intercooler leakage corresponding to the flutter of valve plates and the fluctuation of airflow. This successful diagnosis is achieved by using modulation signal bispectrum magnitude analysis based on compressor vibration signals has taken into consideration both healthy and simulated. It is significant that valve leakage boosts the amplitude of the vibration signals as shown in figure 4. However, it is difficult to separate different compressor fault conditions. In which, MSB results can be based to differentiate the valve leakage and the intercooler leakage from healthy conditions. A higher magnitude can indicate the excessive valve plate's flutters due to the fluctuation of the high-pressure air within the valve system. Furthermore, the differences spread over a wide range of discharge pressures (60psi to 120psi), which is the rated range of the compressor.

REFERENCES

- [1] AlThobiani, F. and A. Ball, *An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks*. Expert Systems with Applications, 2014. **41**(9): p. 4113-4122.
- [2] Ahmed, M., et al. *Fault diagnosis of reciprocating compressors using relevance vector machines with a genetic algorithm based on vibration data*. in *Automation and Computing (ICAC), 2014 20th International Conference on*. 2014. IEEE.
- [3] Elhaj, M.A., *Condition monitoring of reciprocating compressor valves*. 2005, the University of Manchester.
- [4] Haba, U., et al., *Detection and Diagnosis of Compound Faults in a Reciprocating Compressor based on Motor Current Signatures*. 2016.
- [5] Bardou, O. and M. Sidahmed, *Early detection of leakages in the exhaust and discharge systems of reciprocating machines by vibration analysis*. Mechanical Systems and Signal Processing, 1994. **8**(5): p. 551-570.
- [6] Shaeboub, A., et al. *Detection and diagnosis of compound faults in induction motors using electric signals from variable speed drives*. in *Automation and Computing (ICAC), 2016 22nd International Conference on*. 2016. IEEE.
- [7] N. A. Hussein, D.Y.M., E. M. Abdul-Baki, *3-phase Induction Motor Bearing Fault Detection and Isolation using MCSA Technique based on neural network Algorithm*. Journal of Engineering and Development, 2012. **16**(3): p. 15.
- [8] Randall, R.B., *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. 2011: John Wiley & Sons.
- [9] Bloch, H.P. and J.J. Hoefner, *Reciprocating Compressors:: Operation and Maintenance*. 1996: Gulf Professional Publishing.
- [10] Gu, F. and A. Ball, *Use of the Smoothed Pseudo-Wigner-Ville Distribution*. Maintenance & Asset Management Journal, 1995. **10**(2): p. 16-23.
- [11] Wang, Y., M. Liao, and T. Zhao, *Application of the wavelet transform to fault diagnosis in compressor valves*. Zhongguo Jixie Gongcheng/China Mechanical Engineering, 2003. **14**(12): p. 1046-1048.
- [12] Yang, B.-S., et al., *Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines*. Mechanical Systems and Signal Processing, 2005. **19**(2): p. 371-390.

- [13] Lin, Y.-H., H.-C. Wu, and C.-Y. Wu, *Automated condition classification of a reciprocating compressor using time–frequency analysis and an artificial neural network*. Smart materials and structures, 2006. **15**(6): p. 1576.
- [14] Cui, H., et al., *Research on fault diagnosis for reciprocating compressor valve using information entropy and SVM method*. Journal of loss prevention in the process industries, 2009. **22**(6): p. 864-867.
- [15] Yang, B.-S., et al., *Cavitation detection of butterfly valve using support vector machines*. Journal of sound and vibration, 2005. **287**(1): p. 25-43.
- [16] Pichler, K., et al. *Detecting broken reciprocating compressor valves in the PV diagram*. in *Advanced Intelligent Mechatronics (AIM), 2013 IEEE/ASME International Conference on*. 2013. IEEE.
- [17] Elhaj, M., et al., *A combined practical approach to condition monitoring of reciprocating compressors using IAS and dynamic pressure*. Complexity, 2010. **4464**: p. 10062.
- [18] Pichler, K., et al., *Fault detection in reciprocating compressor valves under varying load conditions*. Mechanical Systems and Signal Processing, 2016. **70**: p. 104-119.
- [19] Charles, P., et al., *Detecting the crankshaft torsional vibration of diesel engines for combustion related diagnosis*. Journal of Sound and Vibration, 2009. **321**(3): p. 1171-1185.
- [20] Gu, F., et al., *A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals*. Mechanical Systems and Signal Processing, 2015. **50**: p. 400-413.
- [21] Gu, F., et al., *Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment*. Mechanical Systems and Signal Processing, 2011. **25**(1): p. 360-372.
- [22] Naid, A., et al. *Bispectrum Analysis of Motor Current Signals for Fault Diagnosis of Reciprocating Compressors*. in *Key Engineering Materials*. 2009. Trans Tech Publ.
- [23] Alwodai, A., et al. *Modulation signal bispectrum analysis of motor current signals for stator fault diagnosis*. in *Automation and Computing (ICAC), 2012 18th International Conference on*. 2012. IEEE.
- [24] Hassin, O., et al., *Journal bearing condition monitoring based on the modulation signal bispectrum analysis of vibrations*. 2016, Taylor & Francis.
- [25] Naid, A., F. Gu, and A. Ball, *Fault detection and diagnosis of reciprocating compressors using motor current signature analysis*. 2009, PhD Thesis. University of Huddersfield.

Session 10

Modelling, Analysis and Optimisation 3

Session Chair Dr. Christoph Gerke

A Novel Technique to Reduce Measurement Errors due to Flow –Sensor Interactions in Multi-Sensor Conductivity Probes (*D. Albarzenji, R. Mishra*)

Experimental Model-Based Approach to Integrated Prognostic and Health Management of a Non-Linear Liquid Level System (*A. Al-Khafaji, R. I. Grosvenor*)

Effects of the Fluid Film on the Frequency Response Function of the Structure of Journal Bearings

(*Y. Kang, H Zhang, D. Zhen, Z. Zhi, F Gu*)

Misalignment identification Based on Dynamic Time Warping for Planetary Gearbox

(*Z. Shen, H. Li, D. Zhen, H. Zang, Z. Shi, F Gu*)

A Novel Technique to Reduce Measurement Errors due to Flow–Sensor Interactions in Multi-Sensor Conductivity Probes

Dlir Albarzenji, Rakesh Mishra

School of Computing & Engineering, University of Huddersfield, Queensgate,
Huddersfield, HD1 3DH, UK

ABSTRACT

Multi-sensor conductivity probes rely on multiple sensors intruding into the flow field for the measurement of conductivity variations. This may cause sensors to deflect due to flow-sensor and flow-body interactions. Since this deflection relocates the sensor tips causing inaccuracy in the flow property measurements, many techniques have been used to overcome this issue [1-6]; such as increasing the sensors diameter and reducing the sensors length. However, most of these methods increase the bubble-sensor interactions. In the present work, a novel technique has been developed with the aid of Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) based solvers to reduce the errors that may arise because of the sensor's and probes body's deflections. The developed technique compensates for the errors within the signal processing stage. The CFD model has been validated against experimental data obtained from the literature. Different variables have been investigated to quantify the sensor tip relocation process as a function of pipe diameter, flow velocity and radial probe locations. The results have been presented in the form of mathematical equations using multiple variable regression analyses, and thereafter embedded into the signal processing code.

Keywords: CFD; FEA; Two-phase flow measurement; Four-sensor probe

Corresponding author Dlir Albarzenji (Dlir.Albarzenji@hud.ac.uk)

1. INTRODUCTION

Multiphase flows are integral to many engineering applications; such as boiling, condensing, cavitation, chemical reactions, heat exchanger, oil and gas industries, nuclear plants, etc. Therefore, a wide spectrum of literature is available in which various aspects of multiphase flows have been investigated experimentally and numerically. The dispersed phase flow parameters have been investigated experimentally primarily by using two types of methods; intrusive methods and non-intrusive methods. Among the intrusive methods, various multi-sensor probes have been used successfully for measuring flow properties by various researchers [1-7]. Kataoka et al. [1] have numerically simulated bubbly multiphase flows across two-sensor and four-sensor probes. The authors have investigated the effect of sensor spacing, bubble diameter and bubble-probe contact angle on the accuracy of measurements. In a further development, Kim et al. [2] have developed a four-sensor conductive probe to investigate various flow parameters in a multiphase flow. The studies have also been carried out to quantify the local time average shapes of the bubbles. These results have been benchmarked by processing the images captured using a video camera. Herring and Davies [3] have used a dual-sensor probe to study the dispersed phase local properties in air-water two-phase vertical flow. The authors have reported that the local void fraction profile has remained unaffected by the inlet conditions. Many researchers have been conducted about the effect of probe dimension on the measurement accuracy of the bubble properties. Wu et al. [4] who have concluded that if the axial sensor separation distance (s) to the bubbles diameter (D) ratio is smaller than the maximum relative fluctuation of the bubble velocity the measured bubble velocity may approach infinity value. Therefore, the authors have reported that the accurate results could be achieved only if the range of the axial sensor separation distance locates within 0.5 to 2 of the bubble diameter. Corre and Ishii [5] have numerically simulated the effect of probe geometry on the dispersed phase velocity and interfacial concentration measurements. The authors have suggested a non-dimensional sensor separation parameter

(axial separation divided by bubble diameter) (S/X) in the range of 0.6–1 to achieve accurate velocity measurements in cases where the bubble velocity fluctuations have been relatively low. Shen et al.[6] have experimentally investigated the error sources for optical four-sensor probes in two-phase flow. The authors have attributed the measurement discrepancies to two main sources, namely signal processing and hydrodynamic effect sources. The signal processing source has been related to the threshold value selection, whilst, the hydrodynamic errors have been associated to various phenomena associated with oncoming bubble errors. The probe stiffness has been examined against pressurised air flow; it has been concluded that the optical sensors should be of short length to overcome any deflection.

Based on the literature, it can be concluded that are several factors that affect the accuracy of multi-sensor probes such as the effect of probe location in a pipe and the flow field which causes probe and sensor deflections. Previously, the sensor deflection issue has been dealt with by strengthening the sensors using high strength materials for the sensor body, or by adding sensor support materials around the sensor body or by shortening the sensors length. However, these methods increase the bubble-sensor interactions. Further, none of the researchers tried to overcome this issue by using an embedded code at the signal processing stage instead of the physical treatment. The aim of this research is to estimate and reduce the errors that may occur due to the relocations of the sensors tip because of the deflections of the sensors and the probe body in multi-sensor conductivity probe, using an embedded code at the signal processing stage. The relevant information for the code development is proposed to be obtained from CFD simulations.

1.1. CFD Model and Simulation

Three-dimension computational domains of a vertical straight pipe of 80, 100 and 200mm internal diameter and 2000mm length have been created for simulations. Upward bubbly gas-liquid flows have been modelled using the Eulerian framework of multiphase flow modelling. The continuity and momentum equations for each phase have been solved separately for low volume fractions of approximately 3.8% at three water superficial velocities of 0.76, 3 and 6m/s. The radial velocity and volume fraction distributions of the air have been employed as a criterion for comparing the CFD calculations with the experiments. For the geometry validation, inlet conditions have been assumed to be homogeneous in terms of superficial liquid and gas velocities and volume fractions for both phases in accordance with the experimental setup conditions[8].

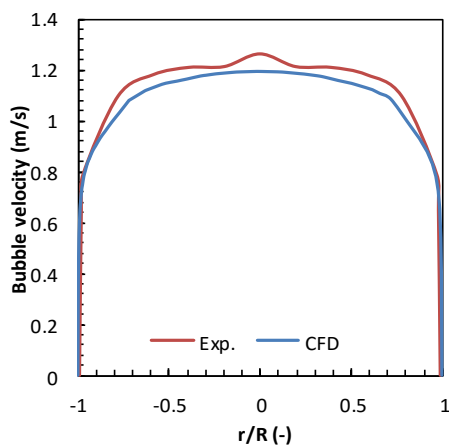


Figure 1. Bubble velocity distribution across the test pipe

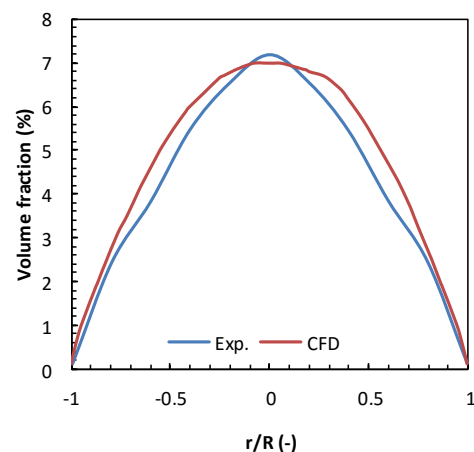


Figure 2. Volume fraction distribution across the test pipe

Figures 1 and 2 depict the ability of the numerical simulation to reproduce the radial volume fraction and the bubble velocity profile and show a good agreement with the experimental measurements. This distribution is mainly influenced by the non-drag forces, which act perpendicular to the flow direction. A lift-force, a wall-force, and a turbulent dispersion-force have been considered in the simulations. For the lift

force, the formulation of Tomiyama model has been used since it can predict the lift force on larger-scale deformable bubbles in the ellipsoidal and spherical cap regimes. As with the Tomiyama drag and wall lubrication models, this model depends on the Eötvös number (Eo).

$$Eo = \frac{g(\rho_L - \rho_G)d_p^2}{\sigma}$$

Where, g is gravity, ρ_L and ρ_G are liquid and gas density respectively and d_p is bubble diameter.

Its main feature is the prediction of the crossover point in bubble size in which particle distortion causes a reversal in the sign of the lift force. The void fraction profile in gas-liquid two-phase flows depends on the drag force to be formed as well as on the non-drag forces [9].

The Shear-Stress Transport (SST) $k - \omega$ model has been selected as a turbulent model. This model has been used to capture the turbulence phenomenon and the flow separation that occurs due to the intrusion of the four-sensor probe. The k - ω SST model is a combination of the k - ω and k - ϵ model, in addition to a shear stress transport model [10]. The typical schematic of the four-sensor probe that had been used by [8] is shown in figure 3.

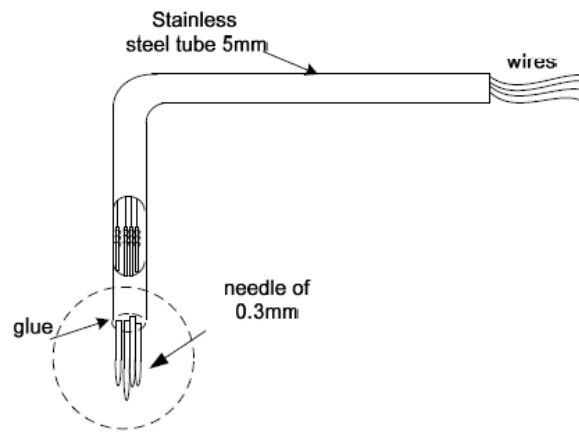


Figure 3. Schematic of four-sensor probe [8].

1.2. FEA simulation

Numerical evaluations have been performed using a commercial finite element code for depicting sensors and probe body displacement effectively. Three-dimensional finite element models have been developed to quantify the maximum displacement that each sensor could have. The static pressure has been transformed from the CFD simulations prediction that have been calculated for two perpendicular planes at each sensor and the probe body dividing each surface into four surfaces, as shown in figure 4.

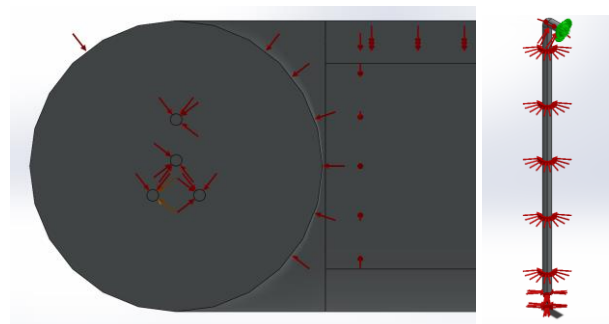


Figure 4. Static pressure distribution around both the four sensors and the probe body

The probe dimension that has been exposed to the fluid flow was 6mm outer diameter and 4.2mm inner diameter. The probe has been assumed to be immersed in the flow domain by 10, 50 and 90mm with 25 and 50mm downstream axial distance.

Chrome stainless steel material has been selected to be the material assigned for both the probe and the four sensors with the specifications, as shown in table 1.

Table 1: The probe and the four-sensor material specification

| Property | Value | Units |
|-------------------------------|-----------|-------------------|
| Elastic Modulus | 2.00E+11 | N/m ² |
| Poisson's Ratio | 0.28 | |
| Shear Modulus | 7.70E+10 | N/m ² |
| Density | 7.85 | Mg/m ³ |
| Tensile Strength | 413613000 | N/m ² |
| Yield Strength | 172339000 | N/m ² |
| Thermal Expansion Coefficient | 1.10E-05 | /K |
| Thermal Conductivity | 18 | W/(m•K) |
| Specific Heat | 460 | J/(kg•K) |

2. RESULTS AND DISCUSSION

In this paper, three parameters have been taken into consideration for investigations namely, the effect of the pipe diameter, the effect of the probe radial distance and the effect of the mixed velocity on the sensor deflection. Whilst the effect of materials that have been used in sensor and probe fabrication, the sensors' length and diameter, and the probe's body diameter have not been taken into consideration and left over for future work.

2.1. The effect of pipe diameter on sensor deflection

One of the important parameter that affects sensor deflection is the use of conductivity probe in different pipe diameters. By using conductivity, probe in larger pipe diameters, with keeping the mixture velocity constant, the sensor deflection decreases, as shown in figure 5. As the pipe diameter increases the pressure drop decreases, which is the reason behind sensor deflection decrease.

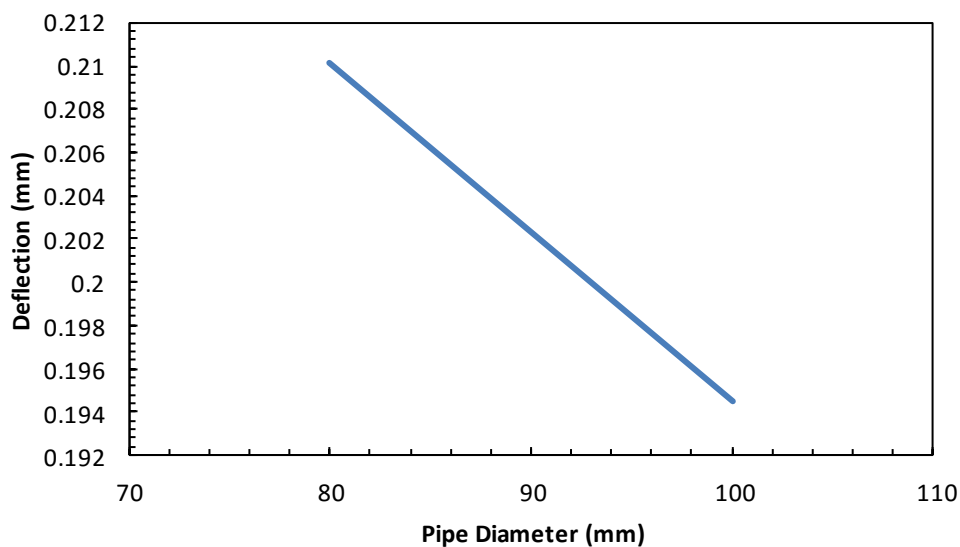


Figure 5. The effect of the change in pipe diameter on the sensor deflection.

2.2. The effect of probe's radial location on the sensor deflection

The effect of the probe location within the pipe on the sensor and the probe body has been considered in this paper at three pipe locations of 10, 50 and 90% of each pipe diameter.

From figure 6, it can be clearly observed that location of probe with respect to pipe walls have a significant effect on the sensor deflection in cases where the flow velocity is relatively low. Maximum deflection can be found when the probe is close to the pipe wall, whilst almost no deflection can be found at the pipe centreline .

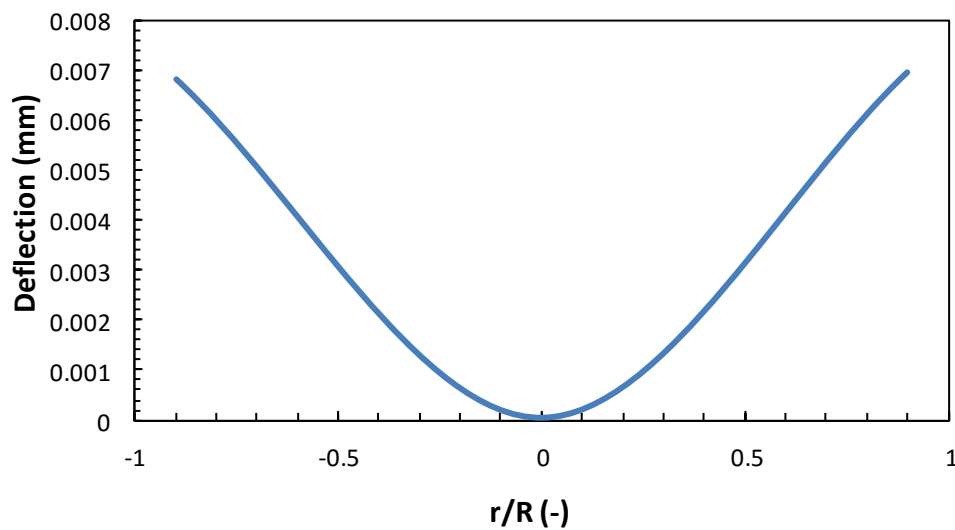


Figure 6. The effect of the change in radial location on the sensor deflection for low flow velocity

Figure 7 depicts the amount of the deflection that is affected by the change in probe's radial location at high velocity of 6m/s. It can be seen that the amount of deflection increases as the probe's stem is exposed to the high flow velocity because maximum probe deflection depends on the perpendicular distance between the applied force and the probe support.

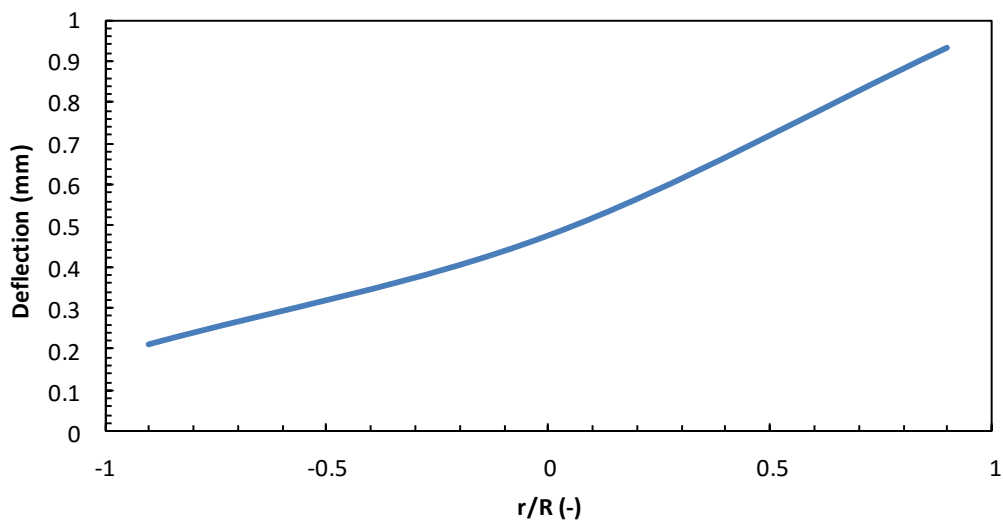


Figure 7. The effect of the change of probes radial location on the sensor deflection

2.3. The effect of the mixed velocity variation on the sensor deflection

Three mixture velocity values of 0.76, 1 and 6m/s have been considered in this study to quantify the effect of the mixed velocity variation on the sensor and probe deflections.

Figure 8, depicts sensor deflection due to the increase in mixture velocity. It can be clearly observed that the increase in flow velocity increases the sensor deflection. This process occurs because of the increase in static pressure that is induced from the frictional shear force at the sensor surfaces due to the high velocity.

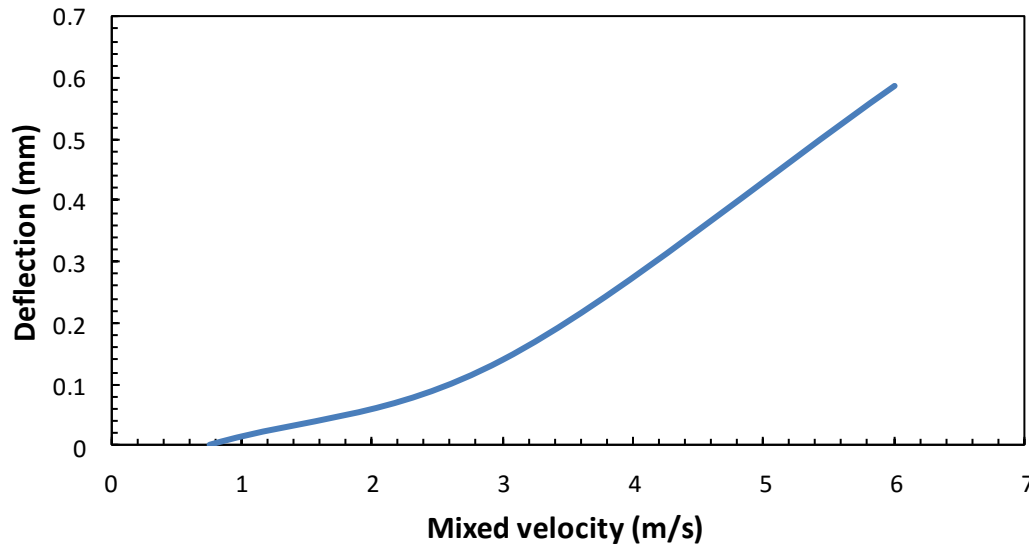


Figure 8. The effect of the change in the mixed velocity on the sensor deflection

2.4. Three-dimensional self-compensation equations

After the investigation of all parameters, the data have been arranged through a regression analysis, which has been used to develop equations represent the sensor deflection in the three directions. For each sensor, three sets of independent deflection equations have been developed.

Sensor 1

$$S1_x = -0.3612 + (1E - 9)Re + (0.0053)r$$

$$S1_Y = -3E - 06 - (8E - 14)Re + (2.42E - 07)r$$

$$S1_Z = 0.0026222 - (4.4457E - 11)Re + (1.1257E - 05)r$$

Sensor 2

$$S2_X = -0.17866225 + (6.64975E - 10)Re + (0.0017776)r$$

$$S2_Y = -0.000411 - (1.5E - 12)Re - (6E - 06)r$$

$$S2_Z = -0.00364 + (4.17E - 11)Re + (1.86E - 05)r$$

Sensor 3

$$S3_X = -0.34975 + (9.96E - 10)Re + (0.0059336)r$$

$$S3_Y = -0.0004 - (1.57E - 12)Re - (5.53E - 06)r$$

$$S2_Z = -0.00375 + (4.09E - 11)Re + (2.2E - 05)r$$

The above equations have been embedded into the signal processing to compensate the experimental data that had been introduced by Pradhan [8]. Figure 9 depicts the effect of the new technique on the bubble velocity distributions. This being the predominant component of flow velocity; the effect of the sensors' deflection on the probe measurement is quite trivial. The maximum discrepancy can be found in the regions where the sensors are close to the pipe wall, whereas almost no compensation can be found at the pipe centreline.

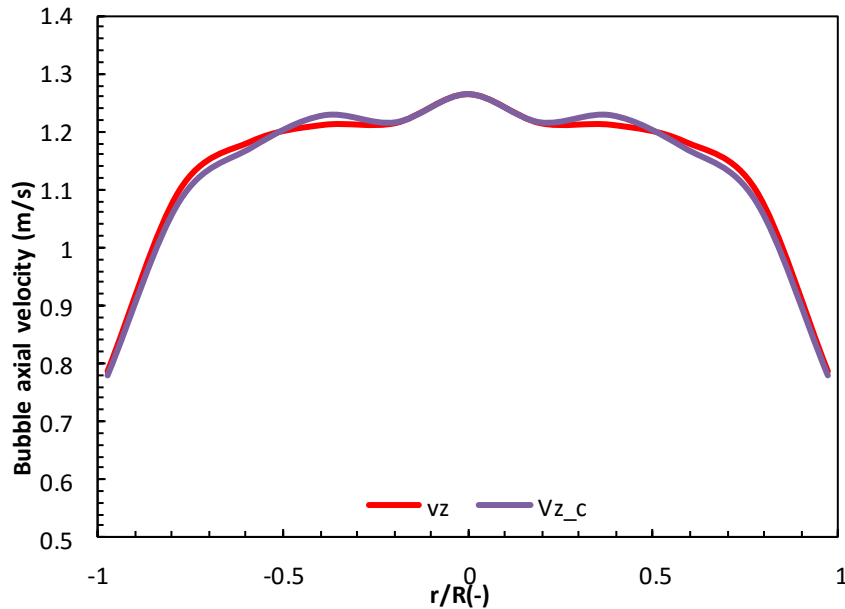


Figure 9. The air axial velocity without correction (v_z) and with correction (v_{z_c})

Figure 10 depicts the effect of the self-compensation technique on the radial velocity. Since the radial velocity is low, the small sensor deflection has high influence on the radial velocity measurement especially in the wall region. However, no effect is found at the pipe centreline.

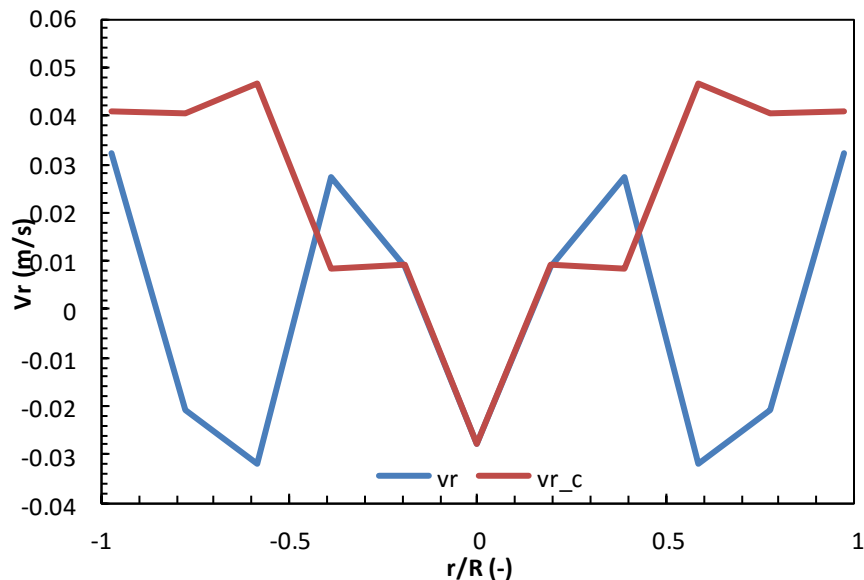


Figure 10. The air radial velocity without correction (v_r) and with correction (v_{r_c}).

Figure 11 depicts the effect of the self-compensation method on the azimuthal air velocity. The effect of the investigated variables on the azimuthal air velocity is trivial in the cases where the mixed velocity is relatively low.

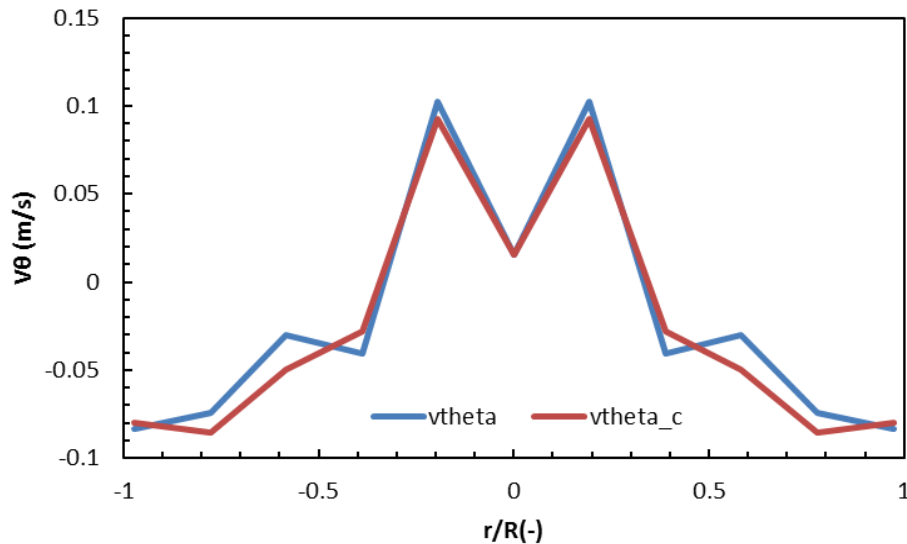


Figure 11. The air azimuthal velocity without correction (v_θ) and with correction (v_{θ_c}).

3. CONCLUSION

Measurement accuracy using multi-sensor conductivity probe depends on a number of factors such as uncertainty in measurement due to sensor and probe deflection, which have been taken into consideration in this paper. The following are the main conclusions:

1. Pipe diameter has a direct effect on the sensor deflection.
2. The position of probe radially affects the sensor deflection rate. At low flow velocity, maximum deflection occurs when the probe gets close to the pipe's wall, whereas at high velocity, the deflection rate increases with the increase of the probe radial intrusiveness.
3. The increase of the mixed flow velocity increases the sensor deflection rate.
4. Using self-compensation technique can effectively reduce errors generated due to sensor and probe body deflections.
5. Combining fluid and structure interaction codes can predict reasonable estimation for the sensor deflection rate.

REFERENCES

- [1] Kataoka, I., M. Ishii, and A. Serizawa, Sensitivity analysis of bubble size and probe geometry on the measurements of interfacial area concentration in gas-liquid two-phase flow. *Nuclear engineering and design*, 1994. 146(1-3): p. 53-70.
- [2] Kim, S., et al., Development of the miniaturized four-sensor conductivity probe and the signal processing scheme. *International Journal of Heat and Mass Transfer*, 2000. 43(22): p. 4101-4118.
- [3] Herringe, R.t. and M. Davis, Structural development of gas-liquid mixture flows. *J. Fluid Mech*, 1976. 73(1): p. 97-123.
- [4] Wu, Q., et al., One-group interfacial area transport in vertical bubbly flow. *International Journal of Heat and Mass Transfer*, 1998. 41(8): p. 1103-1112.
- [5] Le Corre, J.-M. and M. Ishii, Numerical evaluation and correction method for multi-sensor probe measurement techniques in two-phase bubbly flow. *Nuclear engineering and design*, 2002. 216(1): p. 221-238.
- [6] Shen, X., K. Mishima, and H. Nakamura, Error reduction, evaluation and correction for the intrusive optical four-sensor probe measurement in multi-dimensional two-phase flow. *International Journal of Heat and Mass Transfer*, 2008. 51(3-4): p. 882-895.
- [7] Mishra, R., G.P. Lucas, and H. Kieckhefer, A model for obtaining the velocity vectors of spherical droplets in multiphase flows from measurements using an orthogonal four-sensor probe. *Measurement Science and Technology*, 2002. 13(9): p. 1488-1498.
- [8] Pradhan, S., *Measurement of Bubble Velocity Vectors in Bubbly Air Water Multiphase Flow*. 2010, University of Huddersfield.
- [9] Frank, T., J. Shi, and A.D. Burns. Validation of Eulerian multiphase flow models for nuclear safety application. in *Proceeding of the Third International Symposium on Two-Phase Modelling and Experimentation*, Pisa, Italy. 2004.
- [10] Menter, F., et al. The SST turbulence model with improved wall treatment for heat transfer predictions in gas turbines. in *Proceedings of the international gas turbine congress*. 2003.

Experimental Model-Based Approach to Integrated Prognostic and Health Management of a Non-Linear Liquid Level System

Ali A G Al-Khafaji^{1,2}, Roger I Grosvenor¹

¹ School of Engineering, Cardiff University, Queen's Building, Cardiff, CF24 3AA, UK

² Al-Nassiriya Technical Institute, Southern Technical University - Iraq

ABSTRACT

Technology advances have impacted upon monitoring, diagnostic and prognostic activities for increasingly sophisticated industrial systems and their operations. In particular, for integrated mechatronic systems, the facility provided by a dynamic simulation model, as the system experiences deteriorating faults, has been investigated. For informed data-driven prognostic extrapolations, the long-term, time-varying operational profile of the mechatronic system requires recording and analysis. The contribution reported in this paper relates to the simulation, and experimental validation, of a CE 105 coupled-tank liquid level control system. The consideration of a traditional simple system is extended via the inclusion of non-linear elements in the simulation created using the real-time control toolbox within LabVIEW 2014. The simulation is used to accelerate the timescales of the monitoring and tracking system and controller signals for normal and faulty behaviour for several operating condition scenarios. A Sign chart was developed and utilised as a system health monitoring tool. Results are reported and discussed for both leakage and reduced pump performance faults.

Keywords: Sing chart algorithm, non-linear system, System health monitoring; model-based simulation, liquid level system

Corresponding author: Ali A. G. Al-Khafaji (al-khafajiaa@cardiff.ac.uk)

1. INTRODUCTION

A liquid level tank system (LLTS) is used in an extensive variety of industrial process applications, such as water treatment industries, power plants, papermaking and petrochemical industries. Controlling a liquid depth in a tank and consequently, the liquid outflow rate is of crucial importance for mixing reactant process [1].

Reference [2] argued that the control of an LLTS be classified as one of the most challenging benchmark control problems because of its non-linear and non-minimum phase characteristics. An LLTS is commonly controlled by using a conventional Proportional-Integral-Derivative (PID) controller. Such a feedback controller minimises the error between the required demand and the related plant measured variable through regulating the process-controlled inputs. Furthermore, each single element of a PID controller refers to a particular action taken on the error [3].

In considering the dynamic behaviour, and associated simulations, of liquid level control systems the outflow characteristics are usually considered as a primary non-linear fact. However, [4] stated that an LLTS could be regarded as a linear if the free discharge is laminar; $Q = K \cdot h$, where Q is a steady-state liquid free outflow rate (m^3/sec), K is a coefficient (m^2/sec), and h is a steady-state liquid head, (m). When the flow through the outlet restriction is turbulent, the steady-state free outflow rate is $Q = K\sqrt{h}$. Often this characteristic is linearized when the change in the system variables are kept small. For non-linear simulations, a square root characteristic is widely used to model the flow through hydraulic orifices. This may cause numerical problems because the derivative of the flow with respect to the pressure drop tends to infinity when the pressure drop approaches zero [5]. The non-linear is stated to be the characteristic of all real systems [6].

For the approaches in this paper, the model was made as completed as possible and extended to include developing fault conditions. During simulation runs, to represent such prolonged periods the level control system was operated to mimic various demand profile scenarios. These range from a single set point demand for the duration through to a variety of time-varying operating profiles. The simulations were used to evaluate performance and condition by tracking the system and controller signals. The aim is to provide a prior knowledge of the system behaviour for different operational scenarios. Such insights would inform prognostic activities, including remaining useful life (RUL) predictions, potentially with reduced uncertainties. Such a prognostics approach, based on measuring response timings of a pneumatic valve has been reported [7]. In this paper, the importance of a model that describes the nominal and faulty behaviour of a system, and how the latter progresses with the time, was emphasised.

It is stated that whole system prognosis is a difficult task because of some factors such as the uncertainty of the inference processes, fault propagation across subsystems and system complexity in general. Often a limited number of critical components are tracked and used to infer the RUL of the entire system. For example, crack propagation of gears is critical for helicopter transmissions. In petrochemical industries for example, LLTSs may be considered as indicative critical elements. Prognosis approaches are classified as either knowledge – based, data driven or model-based [8].

In condition monitoring applications, it is widely accepted that the control actions will initially mask the early stages of progressive faults. The behaviour of the control signals should then be included as a vital element of the monitoring system. A progressive fault type will not initially affect an observed liquid level, for example with PID control, until a certain level of fault severity is reached [9]. A wide range of application strategies and approaches over several decades have been developed and applied to perform manual, semi-automated, or fully automated system health monitoring of critical systems [10].

In the contribution reported in this paper, a Sign Chart algorithm (SCA) has been created and developed. The SCA provides an ability to diagnose the system faults and monitor the system health. The testing of the approach reported in this paper relates to the simulation and experimental validation of a CE 105 coupled-tank liquid level system. The simulation created used the real-time control toolbox with LabVIEW 2014 software.

2. DESCRIPTION OF THE COUPLED TANK SYSTEM

In this research, a coupled tank apparatus CE 105 was selected to study and simulate for nominal and faulty behaviour. This test rig is designed for teaching process control principles. The system specifications are provided in table 1, and the key features are shown in figure 1. The two tanks are connected to each other, and to a common reservoir through three valves A, B and C. Valve A may be used to control the liquid flow between the two tanks, as it can be seen in figure 1.

Table 1: Specification of a CE105 Coupled Tank System

| | |
|---------------------|---------------------------------------------------------------------------------------------------------|
| Tank 1 | Cross sectional area = 9350 mm^2 |
| Tank 2 | Cross sectional area = 9350 mm^2 |
| Valves A, B, C | 10 mm valve orifice (adjustable), full cross sectional area = 78.5 mm^2 |
| Liquid level sensor | 0 to 10 volt DC output corresponding to 0 to 250 mm as indicates on the front panel water level scales. |
| Flow rate sensor | 0 to 10 volt DC output corresponding to 0 to 4.4 l/min as indicates on the front panel rotameter. |

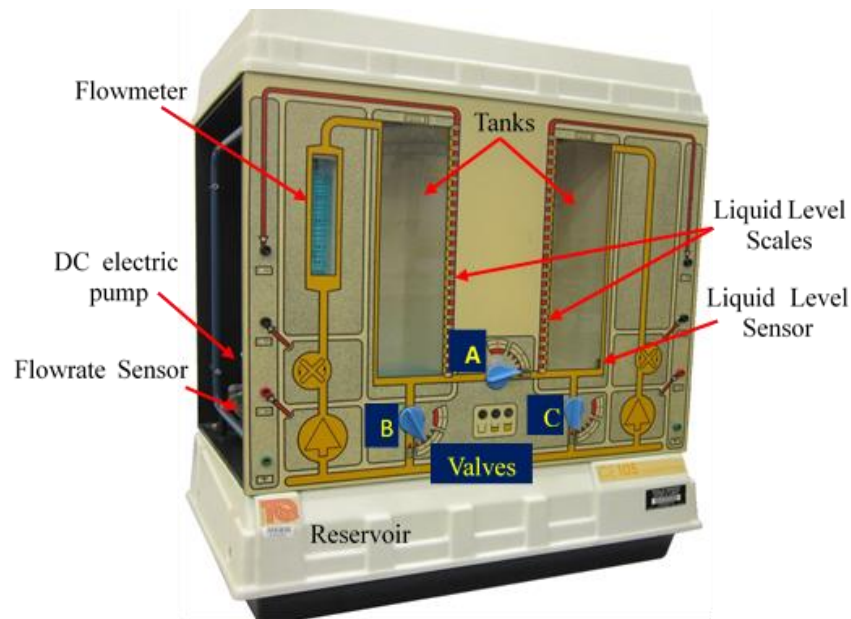


Figure 1. A coupled tank system CE105

Valves B and C are mounted on the individual tank drain pipes, at the height of 46mm measured downward from the bottom of the tank. For this research, valve B was fully closed, and valve C was fully open. Valve A was set to a specific opening proportion to give a selected free outflow rate. Some functions of the usual CE 120 control panel was replaced with a PC interface. LabVIEW control software was interfaced via a USB-NI 6008 DAQ data acquisition board, as shown in figure 2. Pressure-sensing liquid level sensor and a flow rate sensor were interfaced along with an additional flow indicator. The latter is a rotameter installed to provide an inflow rate measurement. Only the left tank was used in the reported testing. A DC motor and pump set was controlled via PID elements of the LabVIEW code.

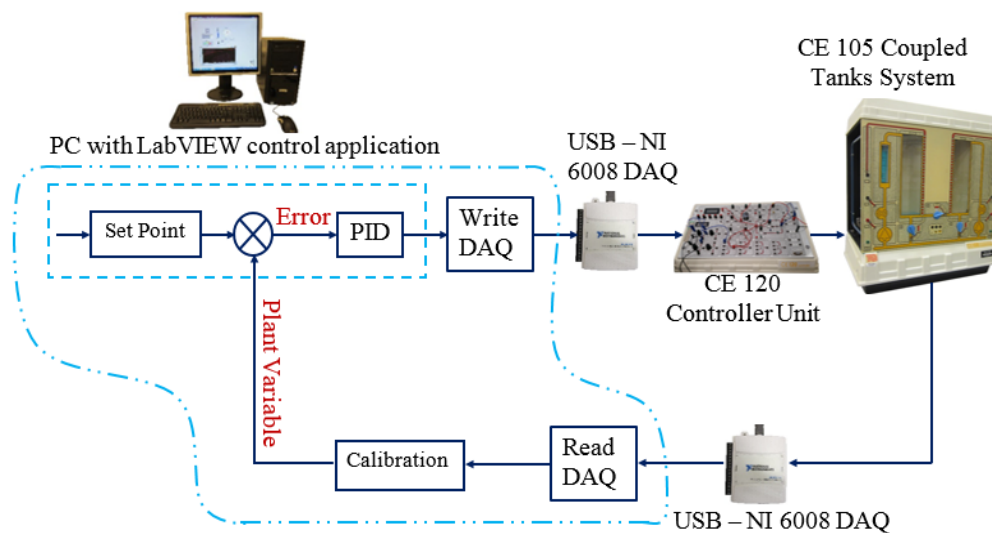


Figure 2. Schematic diagram of the closed loop system

3. CALIBRATION EQUATIONS

The coupled tank apparatus CE 105 was used as a single tank test rig, and open loop experimental tests were performed to estimate the calibration equations of each element of the system. These were used to

populate the LabVIEW simulation, and the results were demonstrated in [9]. The appropriate calibration equations are shown below. Calibration equations of liquid level and flow rate sensors are linear functions companion an offset terms as shown in (1) and (2) respectively.

$$y = 0.0386x + 0.3231 \quad (1)$$

$$y = 2.0954x + 0.2377 \quad (2)$$

Where x is the liquid level (mm); y is the reading of the pressure-sensing liquid level sensor ($volt$) in (1) and the reading of the flow rate sensor ($volt$) in (2).

The Laplace transformation of the DC electric motor water-pump calibration equation used in the simulation was:

$$Q_p = \frac{0.515V_p + 0.0135}{1 + S + 0.29S^2} \quad (3)$$

Where V_p is the pump voltage ($volt$); S is called the (complex) frequency variable sec^{-1} , and Q_p is the pump outflow rate (l/min). The free outflow rate equation is (a non-linear equation), where h is the liquid height (mm):

$$q_o = 0.2383h^{0.514} + 0.003 \quad (4)$$

These equations were used to build a closed loop simulation under LabVIEW 2014 environment to study the system behaviour when some faults may occur.

4. FAULT MODELLING

As a result of ageing or long-term usage, the system behaviour might be deteriorated due to one or more abrupt and/ or incipient faults. The LLTS can be divided into two elements, a high-pressure and a low – pressure side.

- A. The high-pressure side contains the system pump and the tank inlet pipe. A leakage in the tank inlet pipe is assumed to be an abrupt fault. Example incipient faults, which progress slowly with time, were deemed to include pump internal leakage, impeller wear [11]; [12] and bearing wear [12]. Degradation of the pumping efficiency was considered to represent progressive faults and was used in the tests reported.
- B. For a low-pressure side, i.e. tank and drain line, two faults possibilities were considered. An abrupt fault could be represented by a change to the outlet valve setting. Rather manipulating the nominal outflow rate represented a time dependent and slowly progressing fault. For the purpose of this paper, this fault was represented as a percentage of the nominal outflow rate.

The abrupt faults are not be dealt with in this research. Ramp function changes with adjustable slopes were used to represent the progressive faults within the simulation model. Such adjustable function gives an ability to simulate different fault deterioration rates. The simulation is used to accelerate the timescales of the monitoring and tracking the system and controller signals for normal and faulty behaviour for several operating condition scenarios.

5. SIGN CHART METHODOLOGY

Modern industrial processes contain a wide range of elements, sub-systems and sensors. Data from a process under continuous monitoring may become huge, especially when it contains analogue signals sampled at a high rate. Reference [13] investigated a motor current signal from electric control systems for fault diagnosis of centrifugal pumps without installing additional measurement instruments. They concluded that the impeller failure of a centrifugal pump could be diagnosed using remotely measured electric current signals. For a PID-based control system, it might be worth to monitor the controller signal as it is affected by any changing in the required demand, for example. The Sign chart algorithm (SCA) based on a continuous monitoring of the controller signal and the required demand set-point by comparing their voltages at the time (t) with such at the time ($t-1$). Figure 3 shows the principle of a closed-loop controlled system with a Sign chart algorithm. This novel algorithm returns zero if the two voltages are equal to each other, which means the system is at a steady-state and in a healthy condition. The SCA provides an ability to monitor the health of a system, i.e. diagnosis and prognosis. For a liquid level system, the SCA can be utilised for any operating scenario to track the following:

1. Increasing the required demand.
2. Decreasing the required demand.
3. Degradation of the pumping efficiency.
4. Leakage at the low – pressure side.

This algorithm is used to monitor the required demand and a PID output voltage by comparing the current voltage $V_{PID}(t)$ with the previous value $V_{PID}(t-1)$, as follows:

$$Sign = \begin{cases} +1 & \text{if } V_{PID}(t) > V_{PID}(t-1) \\ 0 & \text{if } V_{PID}(t) = V_{PID}(t-1) \\ -1 & \text{if } V_{PID}(t) < V_{PID}(t-1) \end{cases} \quad (5)$$

According to (5), the SCA returns zero if the system at its steady state when there is no change in the controller output voltage. Such algorithm returns other values several times depending on the fault type and its progression pace. Accordingly, it could be easy to diagnose the type of fault and the trend of it regarding the time. The SCA provides a summarised operating profile e.g. how often a fault occurred and for how long it stayed, and how often the liquid level was changed during the operating period.

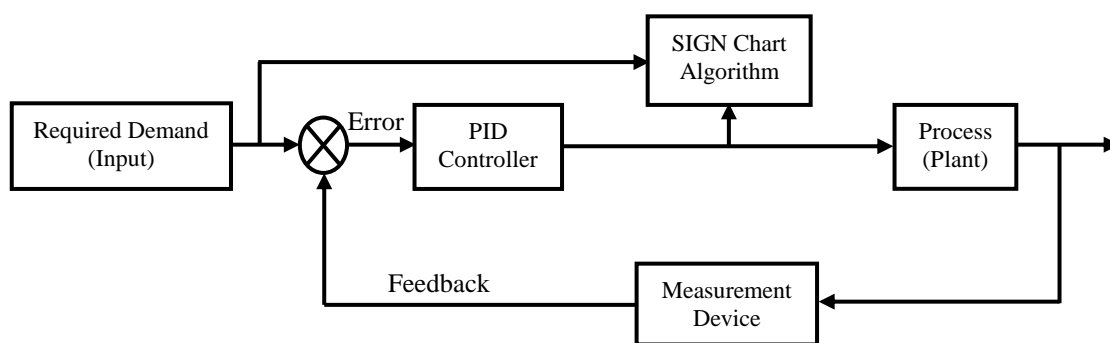


Figure 3. The principle of a feedback control system with a Sign Chart algorithm

6. A CASE STUDY

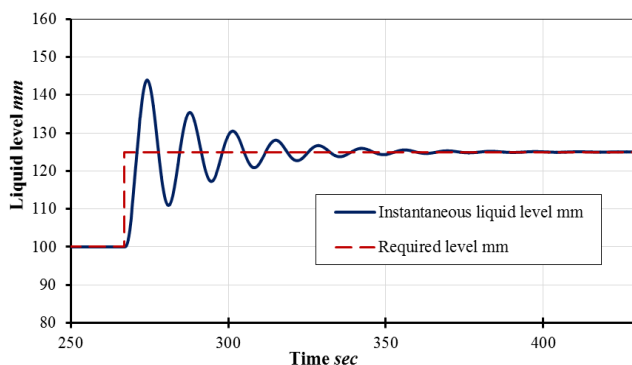
In this study, a PID controller under LabVIEW 2014 environment was used to preserve the desired liquid height and hence the required discharge. Liquid level and outlet valve opening have a direct interactive impact on the free outflow rate. The specifications of the coupled tank apparatus CE 105 in combined with the calibration Equations (1- 4) of the system elements were used to build a closed loop simulation. The PID controller parameters are as shown in table 2. This simulation shows a response as similar as the test rig does at the same system's parameters.

Table 2: The parameters were used for the simulation purpose

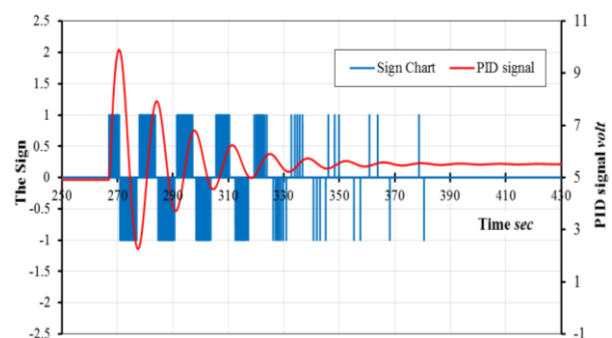
| The liquid level set point | | 100 and 125 mm |
|----------------------------|--------------------------------|-----------------------------|
| PID parameters | Proportional gain (K_c) | 1 |
| | Integral time (T_i , min) | 0.01 min |
| | Derivative time (T_d , min) | 0 min |
| | Output high and low voltage | 10 and 0 volts respectively |

7. RESULTS AND DISCUSSION

1. When the liquid level of a healthy system changed from its steady-state value at (100 mm) to (125 mm), the system response starts fluctuating depending on the setting of the PID parameters. For the purpose of this paper, the PID parameters were at their default values set by a LabVIEW programme, as shown in table 2. As a result of the fluctuation shown in figure 4-a, the SCA returns (+1, 0, -1) values, as presented in figure 4(b). When the system reaches a steady-state, then the Sign value becomes zero.
2. Similarly, if the demand decreased from (125 mm) to (100 mm), The SCA provides an inverse shape of that when the demand was increased, as it can be seen in figure 5.
3. By monitoring the Sign chart, it will be easily realised whether the demand was increased or decreased, at any operating stage.
- 4.

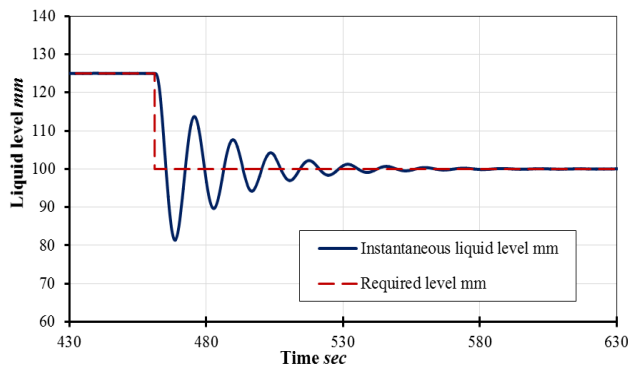


a. Increasing the required liquid level from 100 to 125 mm

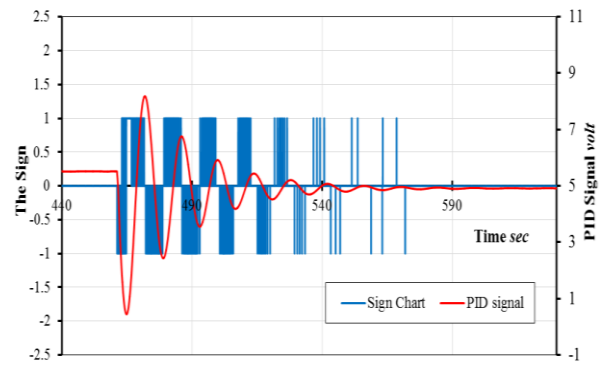


b. The Sign chart

Figure 4. Increasing the required demand



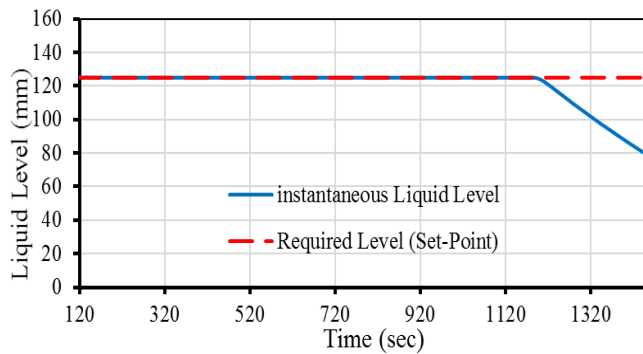
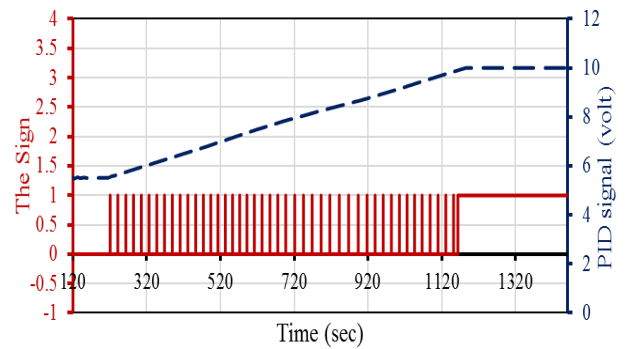
a. Decreasing the required liquid level from 125 to 100 mm



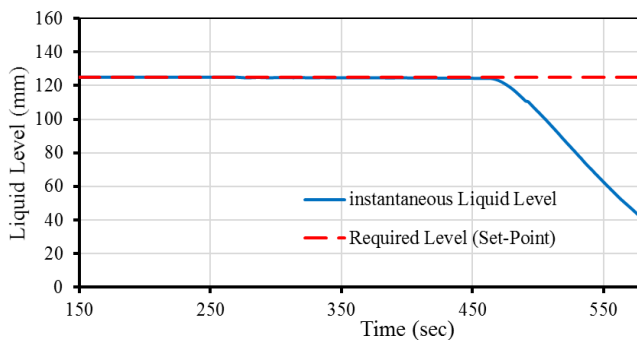
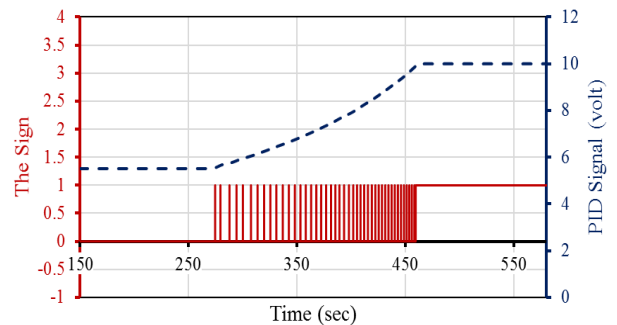
b. The Sign chart

Figure 5. Decreasing the required demand

5. At steady state and when an accelerated leakage, which was mimicked in simulation as a ramp function with a slope equals to (1×10^{-5}) , started at the low-pressure side at the time (230 sec), there was no change apparent in the liquid level. This stability could be attributed to; the PID controller masks the liquid level deficiency by increasing the pump voltage, as it can be seen in figure 6(a). By contrast, the SCA responded immediately by returning (+1) as a result of the PID voltage increment. Fixed spaces between the vertical lines in Figure 6(b) refers to a linear response of the controller as a result of this fault in particular.

a. A leakage fault in the low-pressure side; Slope = 1×10^{-5} 

b. The Sign Chart and The PID signal

Figure 6. A leakage in the low-pressure side; Slope = 1×10^{-5} a. A degradation in the pumping efficiency; Slope = 1×10^{-5} 

b. The Sign Chart and The PID signal

Figure 7. A degradation in the pumping efficiency; Slope = 1×10^{-5}

6. Figure 6(a) shows that the liquid level will not be affected by the fault progression until the PID signal reaches its maximum set value, i.e., (10 volts). Beyond this threshold, the PID controller fails

to mask the fault and hence, the liquid level drops as a result of the fault deterioration. When the PID voltage reaches this saturated-value, the SCA returns (+1) continuously.

7. A degradation of the pumping efficiency, i.e., high-pressure side, makes the system behave differently in compared with the leakage fault at the low-pressure side. As it can be seen in figure 7(b), the spaces between the vertical lines of the sign chart become smaller as time progresses. The system behaviour, as a response to the PID signal due to this fault, is not linear. Meanwhile, the PID response is linear due to leakage fault at the low-pressure side, as shown in figure 6(b).

8. CONCLUSION

The PID controller increases the power supplied to the pump in order to boost the liquid pumping rate to mask any reduction in the required liquid level. This reduction could be a human-made mistake or caused by faults. The SCA converts the PID signal into a simple chart with sufficient details to have an efficient system health monitoring tool. From this chart, it could be easy to monitor the system health and track any changes caused by operator or faults. Traditional monitoring algorithms depend on a massive amount of data collected from the sensors of the system. In contrast of this, the novel Sign chart algorithm provides a simple chart with values (+1, 0, -1) depending only on the controller output voltage, rather than install additional measurement instruments, to monitor the system performance efficiently. As the SCA does not require a massive electronic data-storage medium and because it provides zero at steady-state, it could be easy to track the system operation history.

ACKNOWLEDGMENT

Ali Al-Khafaji would like to thank the Iraqi Establishment of Martyrs and the Iraqi Ministry of Higher Education and Scientific Research (MOHESR) and Southern Technical University – Iraq for their financial support provided.

REFERENCES

- [1] M. Essahafi, "Model Predictive Control (MPC) Applied To Coupled Tank Liquid Level System," *arXiv preprint arXiv:1404.1498*, 2014.
- [2] S. R. Mahapatro, "Control Algorithms for a Two Tank Liquid Level System: An Experimental Study," 2014.
- [3] B. Kumar and R. Dhiman, "Optimization of PID Controller for liquid level tank system using Intelligent Techniques," *Canadian Journal on Electrical and Electronics Engineering*, vol. 2, pp. 531-535, 2011.
- [4] K. Ogata, *Modern Control Engineering*. Upper Saddle River, N. J.: Prentice - Hall, 1997.
- [5] W. Borutzky, B. Barnard, and J. Thoma, "An orifice flow model for laminar and turbulent conditions," *Simulation Modelling Practice and Theory*, vol. 10, pp. 141-152, 2002.
- [6] O. P. V. Himanshu Gupta, "Intelligent Controller for Coupled Tank System," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, 2012.
- [7] M. Daigle, C. S. Kulkarni, and G. Gorospe, "Application of model-based prognostics to a pneumatic valves testbed," in *Aerospace Conference, 2014 IEEE*, 2014, pp. 1-8.
- [8] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, pp. 1156-1168, 2008.
- [9] A. Al-Khafaji and R. Grosvenor "Experimental model-based simulation for health monitoring of a non-linear liquid level system," presented at the ACME-UK 2016 24th Conference on Computational Mechanics, Cardiff University, UK, 2016.
- [10] V. George, L. Frank, R. Michael, H. Andrew, and W. Biqing, "Intelligent fault diagnosis and prognosis for engineering systems," ed: John Wiley & Sons, Inc, 2006.
- [11] G. Biswas and S. Mahadevan, "A hierarchical model-based approach to systems health management," in *Aerospace Conference, 2007 IEEE*, 2007, pp. 1-14.
- [12] M. Daigle and K. Goebel, "Multiple damage progression paths in model-based prognostics," in *Aerospace Conference, 2011 IEEE*, 2011, pp. 1-10.
- [13] X. Tian, G. Feng, Z. Chen, A. Albraik, F. Gu, and A. Ball, "The investigation of motor current signals from a centrifugal pump for fault diagnosis," presented at the 27th International Congress of Condition Monitoring and Diagnostic Engineering, Brisbane Convention and Exhibition Centre, Australia, 2014.

Effects of the Fluid Film on the Frequency Response Function of the Structure of Journal Bearings

Yang Kang¹, Hao Zhang^{1*}, Dong Zhen¹, Zhanqun Shi¹, Fengshou Gu²

¹ School of Mechanical Engineering, Hebei University of Technology, Tianjin, China

² Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK

ABSTRACT

As far as a journal bearing is concerned, the vibration is a common problem. This paper aims at investigating frequency response function (FRF) of a journal bearing under various operating conditions. A dynamics model is developed for the journal bearing considering the effect of the Fluid Film. Modal analysis and harmonic response analysis are conducted. Both natural frequency and FRF of the fluid-solid coupled system are calculated. The distribution of natural frequencies of the journal bearing depends on the stiffness of the fluid film that is also related to the operating conditions. Moreover, a journal bearing test rig is established and the resonances signals are collected from the bearing housing under different radial loads. The experimental results agree well with the trend deduced from the theoretical analysis.

Keywords: hydrodynamic force, fluid-solid coupled system, modal analysis, frequency response function

**Corresponding author: Dr. H. Zhang (zhanghao@hebut.edu.cn)*

1. INTRODUCTION

With the advantages of simple structure, strong carrying capacity, long service life, etc, journal bearings have been widely used in rotating machinery, such as engine crankshaft, turbocharger, ship propulsion shaft and so on. The characteristics of the hydrodynamic force determines the stability of the journal bearings. The hydrodynamic force has an effect on the FRF of the structure of journal bearings. The hydrodynamic force depends on many parameters like clearance, lubricant viscosity and spin speed. Detail derivation of these forces in terms of four stiffness and four damping coefficients are available in Rao [1], Friswell et al [2]. Zhang xiuli [3] provided an efficient method for determining the stiffness coefficients of hydrodynamic plain journal bearings lubricated by water. Zhou hua[4] presented an experimental method to recognize coefficients and established their characteristics under varieties of operating conditions indicated that the linear oil-film dynamic coefficients are sensitive to the excitation force.

Modal analysis is a numerical technique to determine the modal parameter of a system. Modal parameters like modal frequency, modal damping, mode shape and direction of whirl were well explained with the use of complex modal analysis by Joh and Lee [5], Kessler [6], Mesquita et al. Most of the vibration analyses in rotor-dynamics based on natural modes, natural frequencies, and critical speed, reported in many works of literature like Nakra [7], Genta [8] and Friswell et al.

Modal analysis is also used to study the characteristics of the bearing. Y jin [9] built a 3D numerical model of water-lubricated rubber stern bearing and did both free boundary and constraint modal analysis using finite element software ANSYS. The results shows that free modes of water-lubricated rubber stern bearing are correlating with the constraint modes. He XR [10] established the entity model of the bearing-rotor system. The natural frequency and vibration mode of the rotor system are analyzed. Meanwhile, the rule of the natural frequency of the rotor system variation with the oil film stiffness is got by changing the oil film. The results of the analysis shows that supporting stiffness have a great effect on dynamic characteristics of the bearing-rotor system. Wang JF[11] studied the supporting stiffness on the vibration characteristics influence of the axial flow blood pump rotor system. The research shows that in different range of the

supporting stiffness of the bearing, the supporting stiffness basically has no influence on the frequency of the blood pump rotor system, and the modal shape deformation of the blood pump rotor system is tiny.

This paper is to characterise the FRF effected by the fluid film of journal bearings. To present the detail in the accomplishment of the objectives, the paper is organised into five subsections. Following this introduction, a theoretical model is firstly developed for the dynamic characteristics of the journal bearing by taking into account the effects of the fluid film. Then the frequency response curve of the journal bearing are investigated through the simulation results. A journal bearing test rig is established and the effects of stiffness coefficients on the FRF are validated by experimental results.

2. THEORETICAL ANALYSIS

2.1. Dynamical structure of the journal bearing

Due to the micro-irregularities on the surface, the relative motion of journal and bearing coupled with the fluid film produces a fluctuation of the hydrodynamic force. Such force fluctuation yields a broad-banded random excitation which will lead to the mechanical resonances of the supporting system, the housing and the fluid film, as shown in figure 1. The solid part of the supporting system is modelled as a set of mass connected by springs and damping, while the fluid film can be viewed as a constraint with variable dynamic coefficients depending on the operating condition. It is true that the micro-irregularities exist literally on both shaft surface and bearing surface. It is reasonable to analyse it by converting it to the contact between an ideally smooth bearing surface and a rough shaft surface.

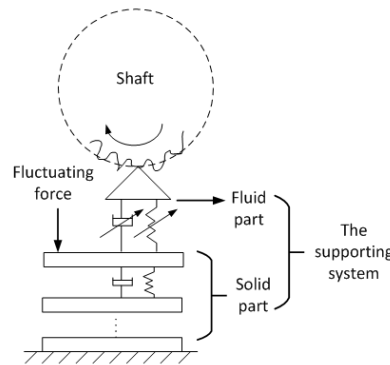


Figure 1. The fluid-solid coupled supporting system of the journal bearing

2.2. Fluid film force

Based on the linear theory, the oil-film force increment of journal bearing is the linear algebraic function of displacements (x, y) and velocities (\dot{x}, \dot{y}) with respect to the journal's static equilibrium position (x_0, y_0) , it can be written as

$$\begin{bmatrix} \Delta \tilde{f}_x \\ \Delta \tilde{f}_y \end{bmatrix} = k_1 \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + d_1 \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} k_{xx} & k_{xy} \\ k_{yx} & k_{yy} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \quad (1)$$

Where $\Delta \tilde{f}_x, \Delta \tilde{f}_y$ are the oil-film force increments in the horizontal and vertical directions. k_1 and d_1 are the linear stiffness and damping coefficient matrices. They have their respective four linear coefficients. It is noted that once multi-sets of displacements (x, y) , velocities (\dot{x}, \dot{y}) and oil-film force increments are obtained from the experiment, the linear oil-film coefficient matrices k_1 and d_1 can be identified.

2.3. Modal analysis of the structure

The dynamic equation of a multi degree of freedom structural system can be described as follows.

$$[M]\ddot{\delta}(t) + [C]\dot{\delta}(t) + [K]\delta(t) = F(t) \quad (2)$$

where $F(t)$ is excitation vector of the load, $\delta(t)$ represents the displacement vector, $[M]$, $[C]$ and $[K]$ are mass, damping and stiffness matrices of the supporting system, respectively.

Without considering the complex modes, the equation for the free vibration is simplified as follows:

$$[M]\ddot{\delta}(t) + [K]\delta(t) = 0 \quad (3)$$

Assuming that the system performing a harmonic vibration yields the following equation:

$$\delta(t) = \varphi \sin(\omega t + \alpha) \quad (4)$$

where φ , ω , α are amplitude, frequency and phase, respectively.

Substituting Eq. (3) into Eq. (2) produces Eq. (4),

$$-\omega^2 [M]\varphi \sin \omega t + [K]\varphi \sin \omega t = 0 \quad (5)$$

To meet the Eq. (4) at any time, the following relation should be satisfied.

$$([K] - \omega^2 [M])\varphi = 0 \quad (6)$$

The constraint on the system can raise the system natural frequencies[11],

$$([K] + [K_r] - \omega^2 [m])\varphi = 0 \quad (7)$$

where $[K_r]$ is stiffness coefficients of the constraint.

The roots of the determinant of the coefficient matrix $\det([K] + [K_r] - \omega^2 [M])\varphi = 0$ are the natural frequencies of the system, and the corresponding eigenvectors are those modes.

3. NUMERICAL ANALYSIS

As shown in figure 2, the 3D model is developed for the self-aligning journal bearing SA35M using UG software. The model is then imported into the workbench for meshing. In order to shorten the time consumed in the calculation later, the structure is simplified in the model development with neglecting chamfering, thread, undercut, fillet, etc.

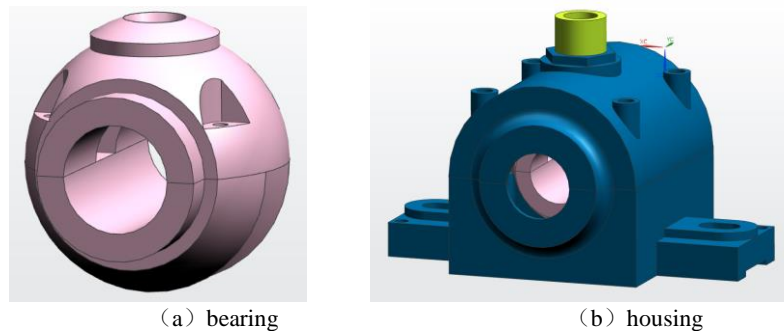


Figure 2. The model of the structure of the journal bearing

The simulation parameters in the workbench are detailed as follows: the material of the bearing housing is cast iron, the elastic modulus is 7.1GPa, the density is 7200kg / m³ and the Poisson's ratio is 0.28; the material of the bearing is copper alloy, the elastic modulus is 1.1GPa, the density is 8300kg / m³ and the Poisson's ratio is 0.34; the material of the oil injection hole is aluminum alloy, the elastic modulus is 7.1GPa, the density is 2770kg / m³, the Poisson's ratio is 0.33.

Figure 3 shows the meshing plot of the structure of the journal bearing. The element size is 8mm. The total number of nodes is 96170 and the number of units is 54394.

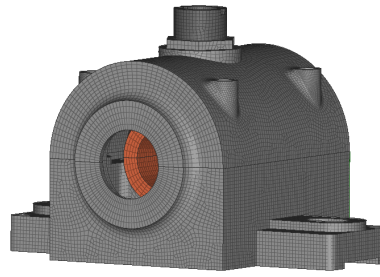


Figure 3. Meshing plot of the structure of the journal bearing

Since the housing is fixed on the base in the real operating condition, an equivalent fixed constraint is applied on the bolt holes in the workbench (figure 4). Two elastic connection constraints are then applied on the bearing to simulate the fluid film and the contact between the bearing and the housing. The latter stiffness coefficient is given to 200N/mm^3 . To simulate the fluctuating hydrodynamic force exerted on the bearing surface, the harmonic force excitation is applied with the value of 50N .

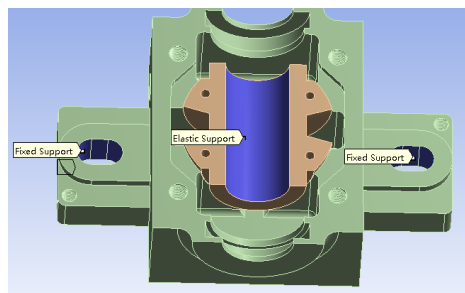


Figure 4. Apply the excitation force to the system

4. RESULTS AND ANALYSIS

4.1. Natural frequencies of the structure

Table 1: Natural frequency of the structure under different stiffness coefficients

| Stiffness value (N/mm^3) | Natural frequency (Hz) | | | | | | | |
|----------------------------------------|------------------------|--------|--------|--------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 200 | 560.83 | 1331.3 | 1895.4 | 2279.3 | 3062.8 | 3394.2 | 3486.7 | 4149.5 |
| 400 | 562.31 | 1351.6 | 1944.5 | 2316.6 | 3438.4 | 3931.0 | 3969.5 | 4342.0 |
| 600 | 562.43 | 1353.4 | 1948.8 | 2319.1 | 3445.6 | 3964.2 | 4038.9 | 4361.1 |
| 800 | 562.50 | 1354.5 | 1951.6 | 2320.7 | 3449.7 | 3984.6 | 4085.2 | 4373.9 |
| 1000 | 563.48 | 1361.3 | 1964.7 | 2327.3 | 3464.7 | 4063.4 | 4300.4 | 4434.5 |

Following the modal analysis, the first eight natural frequencies of the structure of the journal bearing are obtained under different stiffness coefficients, which are listed in table 1. It can be observed from the table that the natural frequencies raise as the stiffness coefficients is increased.

4.2. FRF under different stiffness coefficients

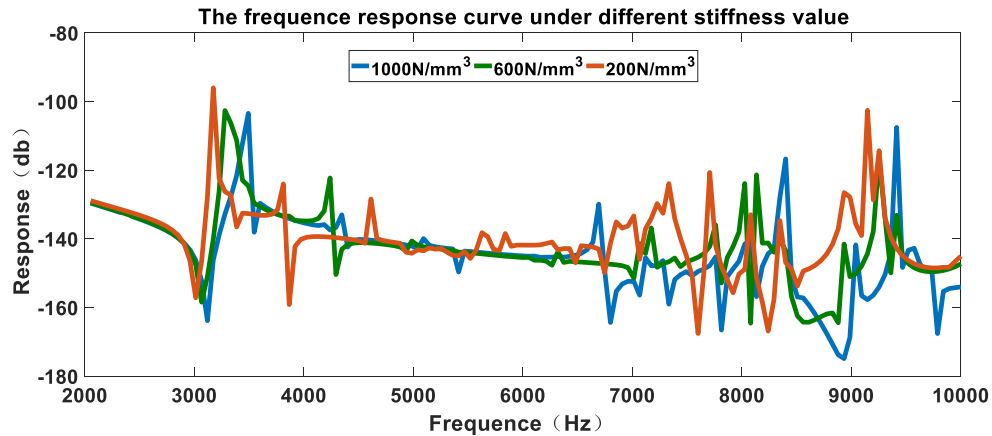


Figure 5. Harmonic response curve under different stiffness

In order to estimate the FRF of the structure, the harmonic response analysis is implemented. A sinusoidal excitation exerted on the inner surface of the bearing and the response is collected on the outer surface of the housing. Figure 5 shows the FRF of the structure from 2kHz to 10kHz under different stiffness coefficients, such as 200N/mm^3 , 600N/mm^3 and 1000N/mm^3 . It can be seen that the frequency response curve are also mainly distributed in two frequency band such as a relatively lower band from 3kHz to 6kHz and a relatively higher band from 7kHz to 10kHz.

5. EXPERIMENTAL VALIDATION

5.1. Test rig description and experiment procedure

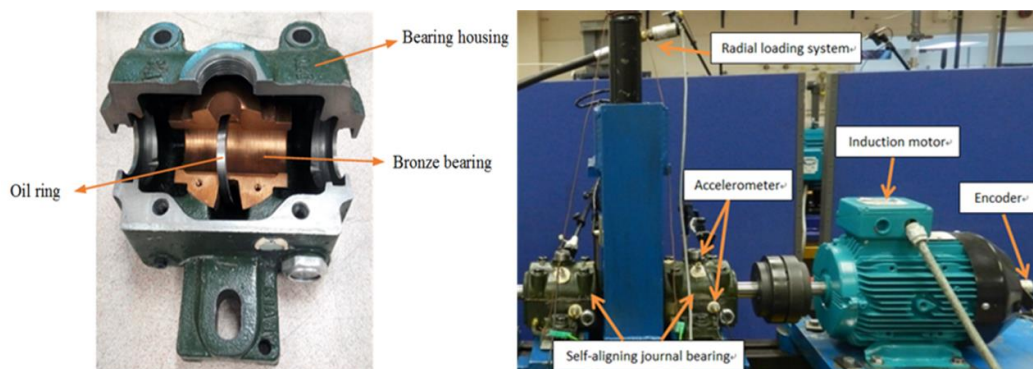


Figure 6. SA35M self-aligning journal bearing and the test rigs

Figure 6 shows the main components of the journal bearing test rig. The shaft is supported on a pair of self-aligning journal bearings. A DC motor drives the shaft spinning and the rotational speed is monitored by an encoder. A hydraulic device is mounted to exert the radial load on the shaft through a rolling bearing. The accelerometers are used to measure the vibration of the bearing housing and a pair of laser sensors are placed vertically for the collection of the shaft trajectory. In addition, the temperature of the lubricant inside the bearing housing is monitored to ensure the bearing working in a healthy condition during the experiment. Measurement of the vibration of the journal bearing is conducted under different radial loads (1bar, 5bar, 10bar, 20bar).

5.2. Harmonic response analysis under different stiffness

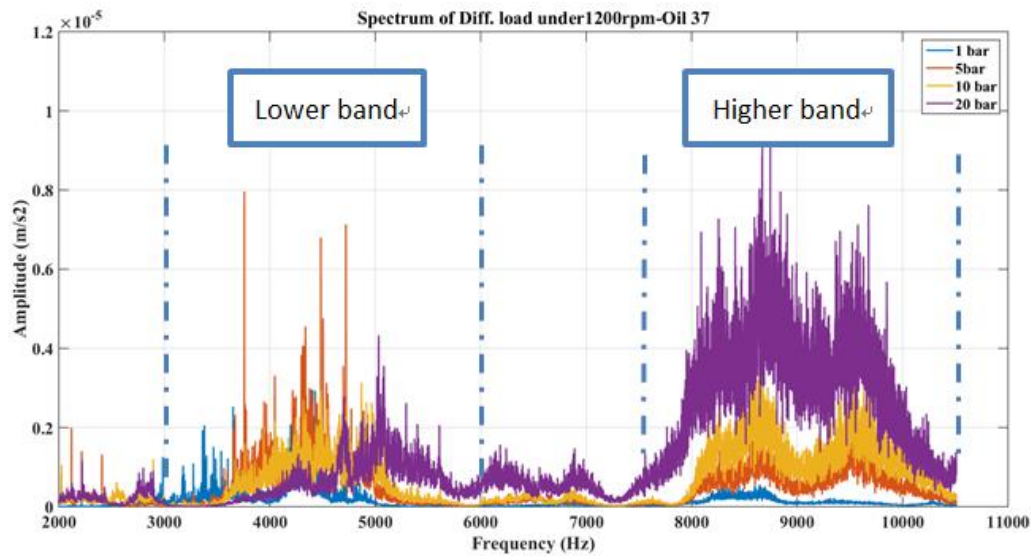


Figure 7. Spectrum of bearing vibration under different radial loads

Figure 7 shows the spectrum of bearing vibration from 2,000Hz to 10,500Hz under different radial load (1bar, 5bar, 10bar, 20bar). It can be seen that vibration components are mainly distributed on two frequency bands. The response amplitude in the higher frequency band increases as the radial load is enlarged, but in the lower frequency band, the vibration components have similar amplitudes. Under a heavier radial load, a larger hydrodynamic force is generated around the shaft to support the external load and thus the amplitudes of the counterforce as well as the fluctuation due to the surface texture become larger. At the same time, the equivalent stiffness coefficient of the supporting system therefore goes up, which contributes to a boost of the system natural frequency. It also can be seen that the crests in the lower frequency band have moved significantly. But in the higher frequency band, the crests have little moved.

6. CONCLUSION

In this paper, the effect of the fluid film on the FRF of the journal bearing is studied theoretically and experimentally. The finite element model of the fluid-solid coupled supporting system is established. The nature frequency and FRF under different stiffness coefficients are calculated. It is found that the increase of the stiffness coefficient of the fluid film will lead a boost to the natural frequency of the supporting system.

The principal stiffness of the fluid film increases under a heavy radial load, which raise the natural frequencies of the journal bearing within a frequency band due to the influence of the contact stiffness between the bearing and housing. The natural frequencies of the supporting system distributed on the lower frequency band are more sensitive to the hydrodynamic lubrication. The operating condition of the journal bearing with a heavier radial load, will give a rise to the natural frequencies of the supporting system in the lower band. The simulation results are validated by the corresponding experimental results.

REFERENCES

- [1] J.S. Rao, Rotor dynamics, New Age International Publishers, New Delhi, 2009.
- [2] M.I. Friswell, J.E.T. Penny, S.D. Garvey, A.W. Lees, Rotor dynamics: modelling and analysis of rotating machines, Cambridge university press, USA, 2010.
- [3] Zhang X, Yin Z, Gao G, et al. Determination of stiffness coefficients of hydrodynamic water-lubricated plain journal bearings[J]. Tribology International, 2015, 85:37-47.

- [4] Zhou H, Zhao S, Xu H, et al. An experimental study on oil-film dynamic coefficients[J]. Tribology International, 2004, 37(3):245-253
- [5] C.Y. Joh, C.W. Lee, Use of dFRFs for diagnosis of asymmetric/anisotropic properties in rotor-bearing system, J. Vib. Acoust. Trans. ASME 118 (1996) 64 – 69.
- [6] C. Kessler, Complex modal analysis of rotating machinery (Ph.D. thesis), University of Cincinnati, United states, 1999, pp. 1 – 106.
- [7] B.C. Nakra, Vibration control in machines and structures using viscoelastic damping, J. Sound Vib. 211 (3) (1998) 449 – 465.
- [8] G. Genta, Dynamics of Rotating Systems, Springer, 2005.
- [9] Jin Y, Lan F. Modal Analysis of Water-Lubricated Rubber Stern Bearing Based on ANSYS[J]. Advanced Materials Research, 2011, 291-294:2111-2114.
- [10] He X R, Fu X J. Research on the bearing-rotor system dynamic characteristics based on ANSYS[J]. Mechanical Research & Application, 2011.
- [11] Wang J F, Sun K, Yun Z. Bearing Stiffness on Modal Influence Analysis of Blood Pump Rotor System[J]. Machinery Design & Manufacture, 2014.

Misalignment Identification Based on Dynamic Time Warping Method for Planetary Gearbox

Zhaoyang Shen¹, Haiyang Li^{1,2}, Dong Zhen^{1*}, Hao Zhang¹, Zhanqun Shi¹ and Fengshou Gu²

¹ School of Mechanical Engineering, Hebei University of Technology, Tianjin, China

² Centre for Efficiency and Performance Engineering, University of Huddersfield, HD1 3DH, UK

ABSTRACT

Condition monitoring (CM) and fault diagnosis of gearbox has gained great attention in industrial applications, and the main techniques widely applied for CM mainly relies on the mechanical vibration. However, the measured vibration signals are normally with low signal-to-noise ratio (SNR) and non-stationary in real practices. It leads to difficult fault detection with high accuracy using vibration signals. This paper proposes a new fault diagnosis method based on dynamic time warping (DTW) for planetary gearbox fault detection using mechanical vibration. DTW is an effective method to align two dissimilarity signals for analysis by stretching and compressing the two dispersed signal sequence in the time domain. So it can be used to reduce noise and extract the fault features as well from the measured vibration signals. Therefore, the characteristics of the measured vibration signals can be improved, and more accuracy detection results can be obtained. The performance evaluation results using both simulated and experimental data shown that the proposed method can enhance the characteristics of vibration signals in the time domain, and the root mean square (RMS) values of the processed signals by DTW were calculated for the misalignment fault identification. It demonstrates that the proposed method can detect the planetary gearbox misalignment faults with different degrees effectively.

Keywords: Planetary gearbox, Misalignment, Dynamic time warping, Root Mean Square.

Corresponding author: d.zhen@hebut.edu.cn

1. INTRODUCTION

Planetary gearbox is widely used in various applications such as the transmission system of wind power generator, helicopter, ship and vehicle due to its structure was compacted as well as transmission ratio of planetary gearbox is high [1]. In real application, the planetary gearbox is usually continuous run for a long time at low speed and heavy load conditions, prone to failure, but its early fault signal is very weak and easy to be submerged in the low-frequency vibration signal and strong background noise, caused that extracting the fault characteristic frequency is very difficult. In addition, because of the compact structure of the planetary gearbox, the space between the accessories and repairs is small. Generally, a shaft misalignment will directly damage a variety of important parts of planetary gearbox, and it might affect the quality of productions, even production stagnation and accidents. Therefore, it is of great significance to study the condition monitoring and the misalignment fault diagnosis of planetary gearbox.

In recent years, many experts and scholars in the field of gearbox CM have found and put forward a large number of research results about the fault diagnosis of planetary gearbox. P.K. Kankar [2,3] explored a methodology for rolling element bearings fault diagnosis using continuous wavelet transform (CWT). Saxena A [4] used a wavelet domain methodology for the analysis and feature extraction of the vibration data from the planetary gear system of military helicopters. Xuefeng Chen [5] presented a novel framework, based on convex optimization, for simultaneously identifying multiple features from superimposed signals. This work adequately exploits the underlying prior information that multiple faults with similar frequency spectrum have different morphological waveforms that can be sparsely represented over the union of redundant dictionaries. Dong Zhen and F Gu [6,7] presented the application of dynamic time warping (DTW) to process the motor current signals for detecting and quantifying common faults in a downstream two-stage reciprocating compressor. F Gu [8-11] proposed a new fault detector based on modulation signal bispectrum analysis (MSB) for bearing fault detection. As MSB results highlight the modulation effects by

suppressing stationary random noise and discrete aperiodic impulses, the detector developed using high magnitudes of MSB can provide optimal frequency bands for fault detection straightforward.

Because of the planetary gearbox has a complex structure, various components of the signal are modulated. In addition, in the process of signal acquisition, signal amplitude, frequency, phase will also produce some errors, especially some data may be lost. Moreover, the fault signal is weak and the signal to noise ratio is low. Moreover, when the planetary gearbox is running at the changing speeds or under the changing loads conventional, many algorithms cannot detected the fault accurately. Therefore, in order to detect the fault signal with high precision, in this paper, a new fault diagnosis method based on dynamic time warping (DTW) is discovered using in vibrations, and it is effective extracted different degrees of misalignment faults of the test planetary gearbox.

The structure of this paper is as follows: The introduction about the research background is explained in Section 1. The theory of dynamic time warping and its significance is declared Section 2. Section 3 provides the vibration signal components of planetary gearbox and the simulation of vibration signal model. Then the method of DTW detection based on the simulation signal, and different degrees of failure were detected. Section 4 introduces the vibration signals are acquired by the experiment of different condition of planetary gearbox. The accuracy of the algorithm is verified by comparing the DTW results with the simulation results. The conclusions and remarks regarding future work are given in Section 5.

2. DYNAMIC TIME WARPING

Dynamic Time Warping (DTW) is an effective algorithm for measuring dissimilarity between time series and has been widely used in various applications [5]. DTW can calculate the sequence of feature vectors and the distance between two time series by extending or shortening the time series. Given two time series named X and Y as Eqs. (1) and (2), and their length are N and M.

$$X = x_1, x_2, x_3, \dots, x_n \quad (1)$$

$$Y = y_1, y_2, y_3, \dots, y_m \quad (2)$$

where x_i and y_j represent the length of the time axis at the point i and j in the series X and Y respectively.

To align the two time series for comparison, a $N \times M$ distance matrix D is built firstly. The element of the matrix D is the distance between the points x_i and y_j which is represented by D_{ij} . Typically, the Euclidean distance is used to calculate the point-to-point distance by Eq. (3).

$$D = \min \sum_{i=1, j=1}^{\max(m, n)} d[x(n), y(m)] \quad (3)$$

where $x(n)$ and $y(m)$ are the two Test template feature sequences . $d[x(n), y(m)]$ is the Euclidean distance of this two time series, and D is the two most similar planning path. But it required heavy computation load. The problem of DTW to be solved is to match the element X and element Y, so that the difference between each pair of matched samples is minimum, and the Euclidean distance is minimized. The alignment path found by DTW is a warping path function which can be defined as:

$$D(n, m) = d(n, m) + \min[D(n-1, m), D(n, m-1), D(n-1, m-1)] \quad (4)$$

where $d(n, m)$ equal to $D(n-1, m-1)$.

Finally, the warping path distance is $D(|X|, |Y|)$. Then, by comparing with the original phase of the model, phase compensation for signal, and the signal modulation effect will more accurate. Then, as long as a fault

model has been establishment, it can extract the fault signal from the vibration signal through applying this model as reference signal..

DTW finds an optimal warping path between the two time series by using dynamic programming to calculate the minimal cumulative distance. There are two time series that have similar overall component shapes, but out of synchronization as shown in figure1 (a). These two time series could alignment by using DTW to stretch or compress the two time series in order to make one resemble another as much as possible as shown in figure1(b).

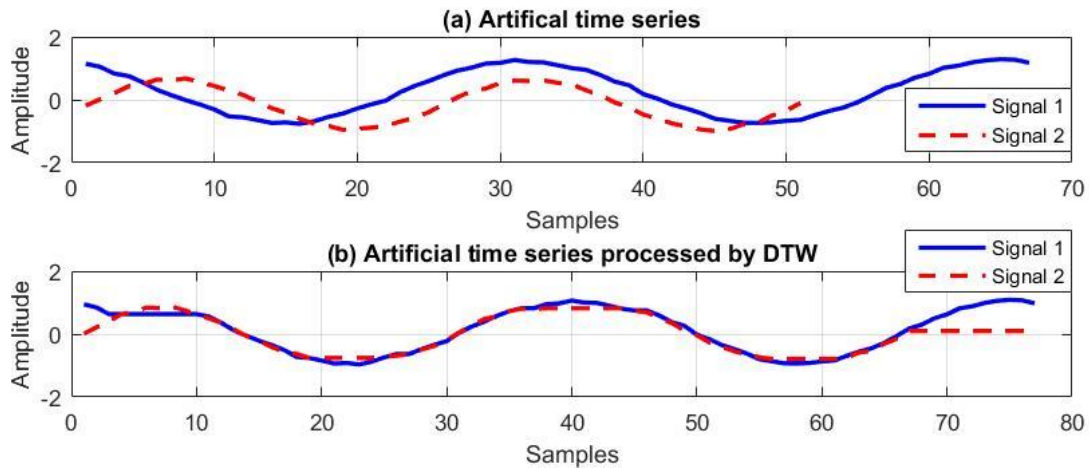


Figure 1. Alignment of two time series based on DTW.

Because of the change of running speeds as well as the loads of planetary gearbox, many fault detection methods cannot be accurately applied. According to figure1, it is not difficult to find out that DTW has the advantage when it confront with the effect by velocity change or by load fluctuation thanks for DTW has the ability to compensating phase for each time series.

3. SIMULATION STUDY

A simulation study was conducted on the vibration signal of planetary gearbox. There are three main components contained in the acquired vibration signal of planetary gearbox. Figure 2(a) shows the amplitude modulation signals of planetary gearbox collected by vibration sensor. Figure 2(b) is a periodic pulse modulated signal that represents a fault in planetary gearbox. Figure 2(c) is the random noise represents the background noise of planetary gearbox. Figure 2(d) is the complete simulation signal that close to the actual acquisition of planetary gearbox vibration signal. Without losing generality, vibration signals from the planetary gearbox with a fault can be expressed as:

$$s(t) = x(t) + f(t) + n(t) \quad (5)$$

where $x(t)$ is collected by vibration sensor with the modulation effect due to the machine structure of planetary gearbox, $f(t)$ is the impulse produced by the fault, and $n(t)$ is the noise which is encountered inevitable in any measurement system.

According to the vibration effect of planetary gearbox, the vibration signal collected by vibration sensor was amplitude modulated. The amplitude modulation of the vibration signal is related to the transmission path, because the transmission path of the vibration signal in the planetary gearbox is still not unified, this article plan the by-pass function as:

$$w(t) = 1 - \sin(2\pi \cdot N \cdot f_c) \quad (6)$$

$$x(t) = w(t) \cdot \sum_N \sin(2\pi \cdot N \cdot f_m \cdot t) \quad (7)$$

where, $w(t)$ is the function of by-pass effect, N is the number of the planetary gears, f_c is Planetary frame rotation frequency, and f_m is the meshing frequency of planetary gearbox.

Once there is a fault in the planetary gearbox, it will produce periodic pulse signal with the operation of the planetary gearbox. The expressions is given below:

$$h(t) = e^{-85t} \cdot \sin(2\pi \cdot 100t) \quad (8)$$

$$g(t) = \sum_k h(t - 0.1k - 0.2) \quad (9)$$

Where, $h(t)$ is a unilateral attenuation signal which modulated by pulse signal and sine signal, and $g(t)$ forcing $h(t)$ appears by cycle.

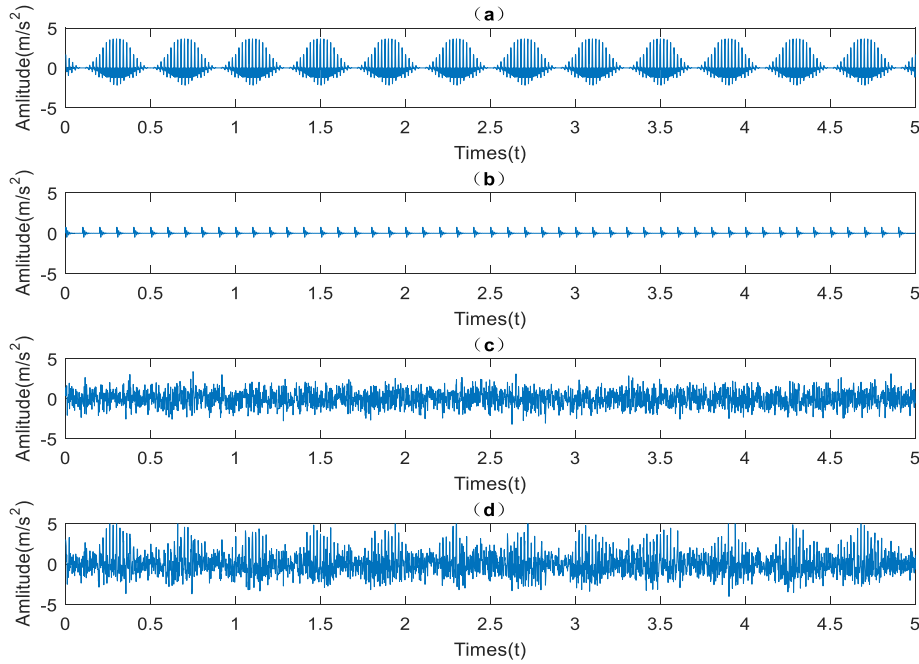


Figure 2. The simulated signals.

In the simulation, the amplitude modulation signal is used as the template, and the synthesized signal to be tested signal. Before the vibration signal collected by sensor is not through dynamic processed, the signal seems is submerged by the noise. Figure 3 (a) shows the synthesized signal and the test signal before the algorithm dynamic warping it. It is obvious that the vibration signal collected by the sensor is stained by the noise signal, so in the time domain signal is not cliffy, and its temporal characteristics are not obvious. Therefore, the fault detection methods are mostly extracted characteristics from the frequency domain of the collected signal.

Dynamic time warping (DTW) allows two time series aligning as much as possible. The fundamental principle of DTW is to stretch and compress two time series in order to make one resemble the other as

possible. Figure 3 (b) shows the synthesized signal and the test signal after the DTW algorithm dynamic processed. It is clear that the processed signal have similar envelope as its template of the amplitude modulation signal. The results show that the DTW can reduce the noise effect.

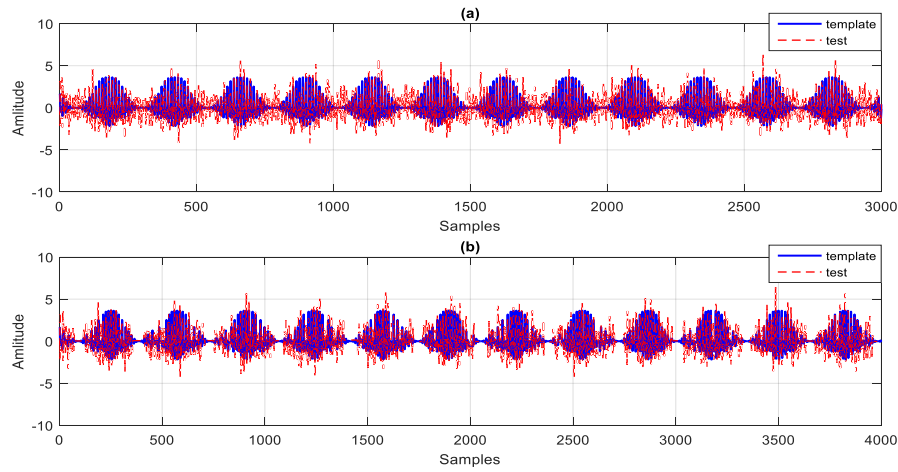


Figure 3. The signal with noise before and after DTW.

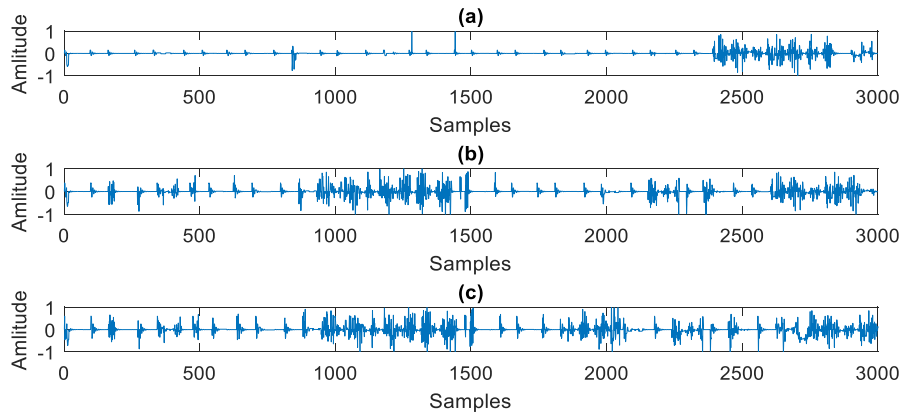


Figure 4. Different fault appearance detected by the DTW.

To get the fault signals, there needs two steps. Firstly, signals of health and faults should be sampled, and taken dynamic processing with the template of amplitude modulation signal to get their residual serials, respectively. Secondly, the residual serial from health is regarded as the template, and other residual serials are obtained and compared with it respectively to get the faults. In order to show the contribution of the DTW to feature extraction in time domain, there is an experiment in which the noise amplitude is at 1.0. Figure 4(a) is a simulation fault signal at 0.2 amplitude detected by the DTW, Figure 4(b) is a simulation fault signal at 0.5 amplitude detected by the DTW, Figure 4(c) is a simulation fault signal at 0.8 amplitude detected by the DTW.

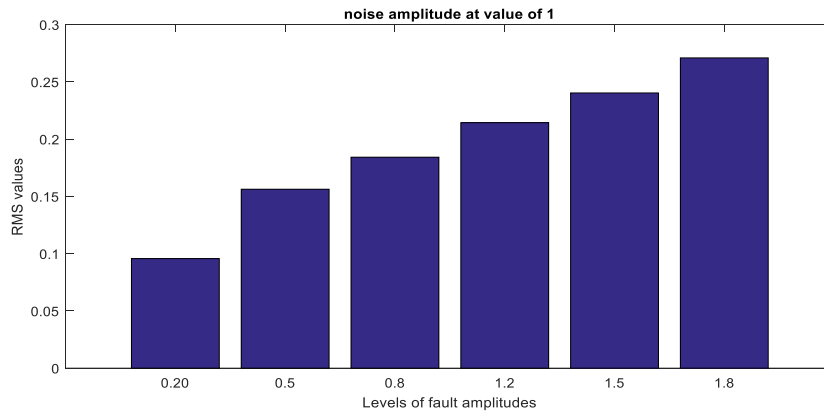


Figure5. The RMS values of different faults detected by the DTW.

In order to verify the characteristics of the vibration signal of the planetary gearbox processed by DTW algorithm, there the Root Mean Square (RMS) values were calculated for different amplitudes of fault signals below and higher than the application of DTW in the same noise background are shown in figure 5. The result shows that as long as planetary gearbox running in the same condition, the higher the amplitude of the impulse vibration of the fault did, the higher the RMS value of it arrived. This phenomenon shows that the detection result by the misalignment detector based on the DTW can avoid the effect caused by the noise signal.

4. MISALIGNMENT FAULT IN A PLANETARY GEARBOX

To verify the conclusion of simulation experiment and the effectiveness of DTW algorithm to boost up the features of signal, experiments are investigated on a planetary gearbox test system, and from which acquired the vibration signals. In a set of experiments, Varying degrees of misalignment for the planetary gearbox were designed. The fault is achieved by placing different thickness gaskets on the base of the experimental platform. The faults are at 0.4mm misalignment rate, at 0.7mm misalignment rate, at 1.0mm misalignment rate respectively.

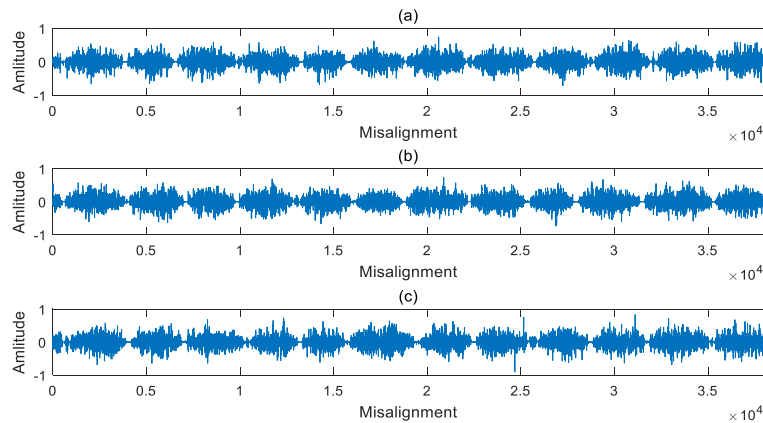


Figure 6. The residual signals of different misalignment after DTW.

The vibration signals acquired from the planetary gearbox are warping compared with the template by DTW. Then, dynamic processing Figure the residual signals of fault with residual signals of healthy as template. Figure 6(a) shows the residual signal of misalignment fault at 0.4mm rate, Figure 6(b) shows the residual signal of misalignment fault at 0.7mm rate, and figure 6(c) shows the residual signal of misalignment fault at 1.0mm rate. As can be seen from the figure 6, they have the same envelope.

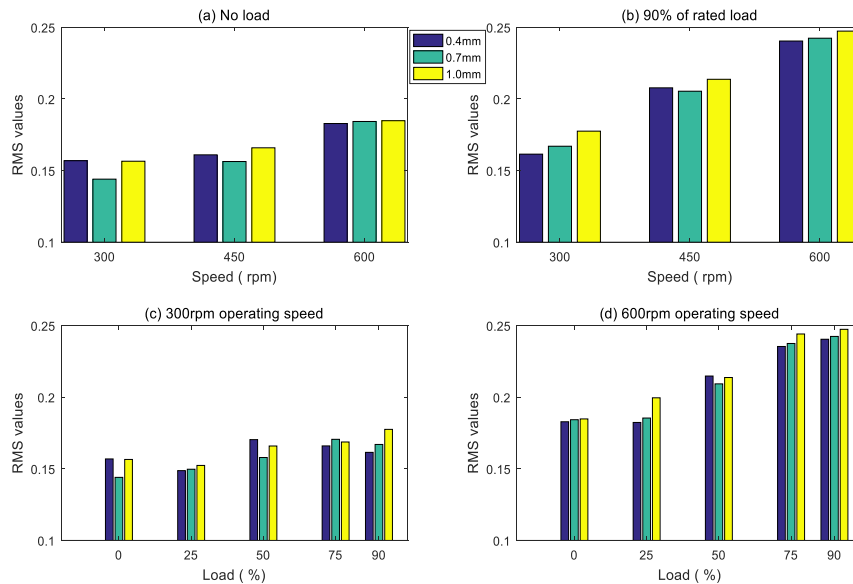


Figure 7. The RMS values of misalignment in different condition.

Figure 7(a) shows the RMS of different misalignments influenced by different rotating speeds when there is no load. Figure 7(b) shows the RMS of different misalignments influenced by different rotating speeds when the load is at 90% of the rated load. Figure 7(c) shows the RMS of different misalignment influenced by different load when the operating speed is set at 300 rpm. Figure 7(d) the RMS of different misalignment influenced by different load when the operating speed is set at 600 rpm.

By analyzing the results in figure 7, it can be found that the RMS value is increased along with the operating speed increasing obviously as shown in figure 7 (a) and (b). The RMS value is increasing along with the increasing of the planetary gearbox load as shown in figure 7 (c) and (d). And generally, the RMS value is increased along with the degree of misalignment rising at the same condition. However the changing of RMS values caused by the loads is more sensitive than it caused by the operating speeds. But the results cannot present well when it at low operating speed, and different degree of failures is not obvious at low loads.

5. CONCLUSIONS

The misalignment detector based on the DTW method can identify the misalignment faults at different degrees effectively under varies operating conditions in the time domain. It is suitable for the fault diagnosis of planetary gearbox and the noise has no effect on the performance of misalignment detection. The simulation and experimental results shown that the RMS value is increased along with the degree of misalignment rising at the same condition. In addition, experimental results shown the changing of RMS values caused by the load is more sensitive than it caused by the operating speed. It presents that the RMS values of the residual signals processed by DTW can be used to identify the misalignment faults of planetary gearbox under different operating conditions.

ACKNOWLEDGMENTS

This research was financially supported by the National Natural Science Foundation of China (Nos. 51605133), the Tianjin Natural Science Foundation of China (Grant nos. 14JCYBJC42100), and the international science and technology cooperation project of Hebei province (Grant nos. 17394303D).

REFERENCES

- [1] LEI Ya-guo, HE Zheng-jia, LIN Jing, et al. Research advances of fault diagnosis technique for planetary gearboxes [J]. *Journal of Mechanical Engineering*, 2011, 47(19):59 — 67.
- [2] P.K.Kankar, S.C.Sharma, S.P.Harsha. Fault diagnosis of ball bearings using continuous wavelet transform [J]. *Applied Soft Computing*.2011, 11(2): 2300-2312.
- [3] PK Kanka, SC Sharma, SP Harsha. Fault diagnosis of rolling element bearing using cyclic autocorrelation and wavelet transform [J]. *Neurocomputing*, 2013, 110(8):9-17.
- [4] Saxena A, Wu Bqing, VACHTSEVANOS G. A methodology for analyzing vibration data from planetary gear systems using complex Morlet wavelets [C]. *American Control Conference*, June 8-10, 2005, USA. 2005:4730-4735.
- [5] Z Du, X Chen, H Zhang, R Yan. Sparse Feature Identification Based on Union of Redundant Dictionary for Wind Turbine Gearbox Fault Diagnosis [J]. *IEEE Transactions on Industrial Electronics*, 2015, 62(10):6594-6605.
- [6] D Zhen, HL Zhao, F Gu, AD Ball. Phase-compensation-based dynamic time warping for fault diagnosis using the motor current signal [J]. *Measurement Science & Technology*, 2012, 23(5):462-467.
- [7] D Zhen, T Wang, F Gu, AD Ball. Fault diagnosis of motor drives using stator current signal analysis based on dynamic time warping [J]. *Mechanical Systems & Signal Processing*, 2013, 34(1-2):191-202.
- [8] F Gu, Y Shao, N Hu, A Naid, A D Ball. Electrical motor current signal analysis using a modified bispectrum for fault diagnosis of downstream mechanical equipment [J]. *Mechanical Systems and Signal Processing*, 2011, 25(1):360-372.
- [9] A. Alwodai, T Wang, Z Chen, F Gu, R Cattley, A. Ball. A Study of Motor Bearing Fault Diagnosis using Modulation Signal Bispectrum Analysis of Motor Current Signals. *JSIP*, 2013, 04(03):72-79.
- [10] F Gu, T Wang, A Alwodai a, X.Tian, Y Shao, A D Ball. A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals [J]. *Mechanical Systems & Signal Processing*, 2015, 50:400-413.
- [11] X Tian, GM Abdallaa, I Rehab, F Gu. Diagnosis of combination faults in a planetary gearbox using a modulation signal bispectrum based sideband estimator [C]. *International Conference on Automation & Computing*, UK, 2015.

Session 11

End User Applications and Maintenance in Industry 2

Session Chair Dr. Jonas Albers

Effect of Manufacturing Method of a Centrifugal Fan Hub on its Heat Dissipation Characteristics

(J. Swinton, T. Eshaafi, T. Asim, J. Irons, R. Mishra)

Abrasive Spur Gear Wear Prediction Model: Utilization of the Statistical Design of Experiment

(S. Raadhui, K Wangveera)

Degradation-Based Preventive Maintenance Policy for Railway Transport Systems

(F. Dinmohammadi, B. Alkali, M. Shafiee)

Condition Based Maintenance Decision Making: a Practical Approach for Marine Vessels

(L. G. Huber, S. Kunz, M. Dettling)

Effect of Manufacturing Method of a Centrifugal Fan Hub on its Heat Dissipation Characteristics

James Swinton¹, Taher Eshaafi², Taimoor Asim^{*3}, John Irons⁴, Rakesh Mishra⁵

^{1,4}Halifax Fans Ltd, Brookfoot Business Park, Elland Road, Brighouse, HD6 2SD, UK

^{2,3,5}School of Computing & Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK

JamesSwinton@halifax-fan.com¹, Taher.Eshaafi@hud.ac.uk², t.asim@hud.ac.uk³, JohnIrons@halifax-fan.com⁴, r.mishra@hud.ac.uk⁵

ABSTRACT

As the process temperature of a fan system increases, the amount of heat that gets transmitted to the bearings and/or motor increases. If this is not accounted for, it can lead to catastrophic failure. The main heat conduction path is through the shaft, and certain mechanisms must be considered when looking for new solutions. These include; how heat is transmitted through the shaft or increasing the thermal resistance of the shaft, and dissipating heat as it is conducted through the shaft. These aspects must always be considered in addition to the impact of the manufacturing complexity. In the present study, an existing heat dissipation arrangement is reviewed and replaced by a new design which reduces the time taken to machine the part, and ultimately the overall cost of the product. Computation Fluid Dynamics (CFD) based techniques have been used to numerically simulate the designs under operating conditions, and the resulting heat transfer through the shaft compared with respect to the heat dissipation properties. The results demonstrate that although the new design is less effective at dissipating heat, it provides a substantial cost reduction compared to the existing design, while substantially reducing the impact of the design on various aspects of production.

Keywords: CFD, Heat Transfer, Centrifugal Fan

Corresponding author: J. Swinton (JamesSwinton@halifax-fan.com)

1. INTRODUCTION

The application of fans in high temperature environments requires the control of heat transfer through its components, in particular through the shaft as this connects the impeller to the prime mover, which moves the high temperature gas through the system, to the sensitive areas such as bearings that lie within a motor. Heat transfer can create problems for equipment operating at high temperatures, if the phenomenon is not understood and equipment is designed incorrectly the outcomes can be less than desirable. Currently a special coupling is utilised for high temperature applications, in order to dissipate heat. The coupling connects the impeller to the motor shaft, the impeller would be subjected to the high temperatures whilst the motor would remain below 70°C. According to Fourier's law heat transfer will take place through the shaft, to what extent depends on material and geometric properties. Should the temperature be too great at the point along the shaft where the motor bearings sit, permeant damage will occur and result in bearing failure. The current design is quite labour intensive to produce, because of this a new design has been created that can be easily machined, thus reducing the overall time spent manufacturing, which in doing so reduces the overall cost of the product. CFD analysis has been carried out at 300°C, the temperature at which the hubs would be exposed to, to analyse if the new design has the same heat dissipation properties as the baseline model.

Researchers have analysed how changes to the shaft speed, fin arrangement and fin thickness affects the transfer of heat through a system and away from a shaft. Zainullin.et al. [1] carried out an experimental investigations of a furnace fan to determine how changing fan speed and enclosing the shaft from its surroundings affects the heat transfer from the shaft. The results indicate that the heat transfer coefficient were 40-60% higher for a shaft that was open and rotating compared to an enclosed stationary shaft. Aziz.et al. [2] looked at solutions for rotating radial fins and how heat is lost to their surroundings, the study analysed theoretically using homotopy analysis to determine how fin thickness affected the results. The

results gained when compared directly to numerical solutions to show the accuracy of the theoretical work. Watel.et al. [3] studied the effect of fin cooling experimentally using infrared thermography. The study explores how the rotational speed and fin spacing affect the heat exchanged in a finned tube. Using the results the mean value for the convective heat transfer from the rotating finned tube and the heat transfer coefficient from the cooling process could be calculated. Xie.et al. [4] experimentally studied a rotating and a stationary heat pipe in a condenser to investigate how the temperature in the pipe changes due to the rotation. By measuring the temperature at five points along the shaft using infrared thermocouples, it was possible to determine that when using the rotating shaft the temperature difference between the condenser and the shaft was reduced to almost zero, opposed to the stationary shaft whose temperature was much higher. Mori.et al. [5] present a CFD analysis of heat transfer on rotating blades. The boundary conditions represented real world characteristics that the geometry would be experience in practice. The results were verified experimentally using infrared thermography to study the temperature distribution. The method developed from the work can be used in other rotating machinery examples where accessing the object to gain readings is problematic.

From the review of the published literature there is good knowledge within the field, using apparatus to dissipate heat away from rotating shafts. The majority of examples found use fins in various arrangements and have calculations that are able to support their findings. Where there are potential areas lacking in knowledge would be the direct application of heat dissipation products for use in fan environments. In the present study, numerical simulations have been carried out investigating the differences in the heat transfer of a rotating component, that arise from modifying is geometry.

2. NUMERICAL MODELLING

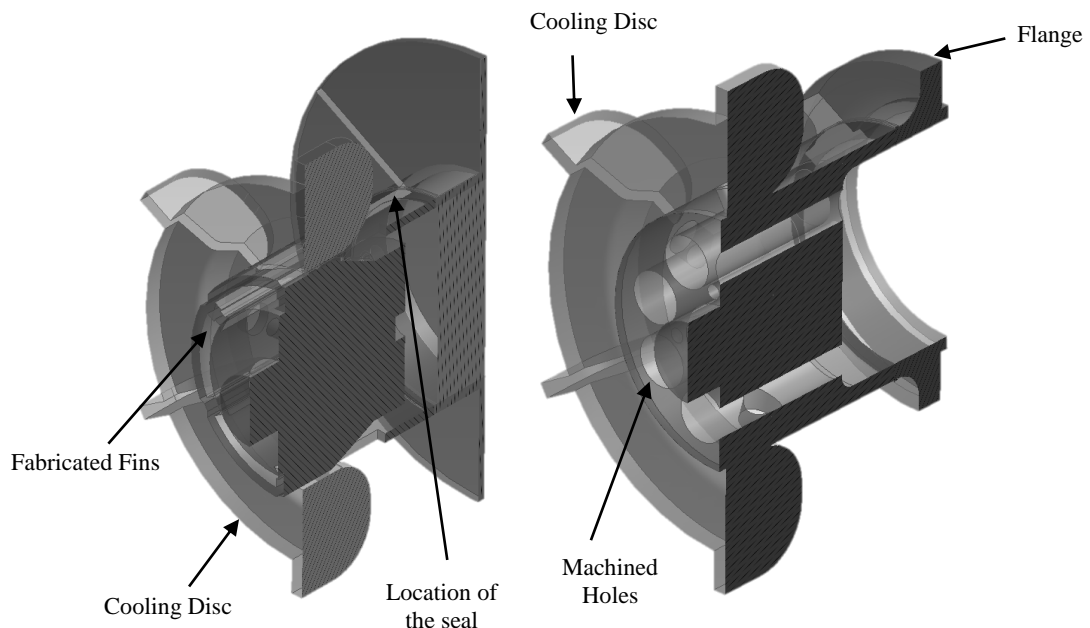


Figure 1. Cross sectional views showing the two geometries

For the numerical solutions, two models using the old and new hub designs made up, these formed part of a system that included; the shaft with the hub attached, a small gap to simulate the seal between the fan casing and the shaft and an area representing the bearing unit terminating the shaft. The shaft has a diameter of 28mm and the point at which it reaches the bearing unit is 15mm from the hub body. When modelling the hubs the elements kept the same for both models were the overall length, diameter, materials (the hub body is carbon steel and the cooling disc aluminium) and the cooling disc position. The cooling disc is bolted onto the hub and provides a key role in dissipating heat away from the shaft; this is due mainly to its finned

design. To simulate the cooling disc and shaft rotation a moving reference frame (MRF) approach has been applied.

Figure 1 show both hub designs, the old hub with the internal fin sections within the main body show the complexity of the design as these have to be fabricated individually to the body. It also shows the new hub design with its simplified construction as a single machined part.

2.2 Meshing of the Flow Domain

For the simulation a hybrid mesh has been used, utilised both hexagonal and polyhedral elements. The domain consisted of 5.5 million elements in total, with 1.1 million for the solids zone and 2.2 million for the fluid zone. It has been noticed that the mesh considered in the present study can be effectively used to capture the complex flow phenomena, and the associated heat transfer, with reasonable accuracy.

2.3 Boundary Conditions

The boundary conditions for the simulation are summarised in table 1.

Table 1: Boundary Conditions

| Item | Boundary Type | Value |
|-------------------|------------------|---------|
| Flange | Heat | 300°C |
| Seal | Pressure (Total) | 2000Pa |
| Cooling Disc | MRF Zone | 1500rpm |
| Ambient of Domain | Heat | 26.85°C |

3. RESULTS AND ANALYSIS

The heat dissipation coupling models, both the old and the new design have been analysed at a temperature of 300°C. Figures 2 and 3 show the results from the heat transfer analysis and the differences between the models are highlighted and discussed. Figure 2 shows a cross section of the old hub model, the 300°C section can clearly be seen on the flanged section of the hub. The temperature distribution displays how the geometry of the hub affects the temperature through the different parts of the model.

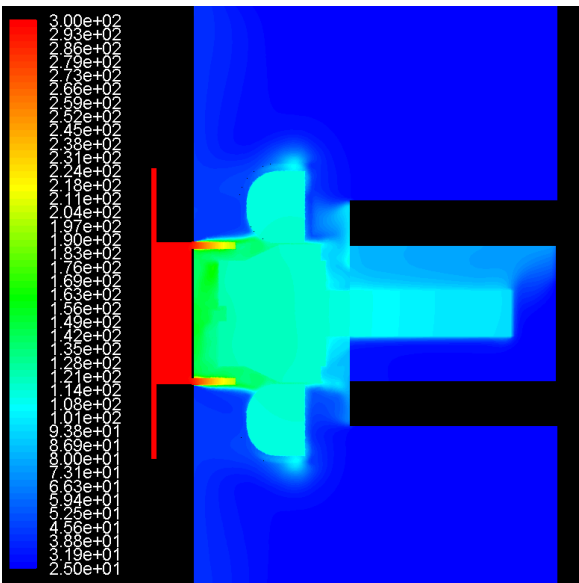


Figure 2. section view of the old hub design

Figure 3 shows a section view of the new machined hub, the temperature of 300°C has again been set to the flange section, on this model the flange is smaller. By drawing a direct comparison between the two models it is clear to see that after the seal section of the new hub model the overall temperature is increased.

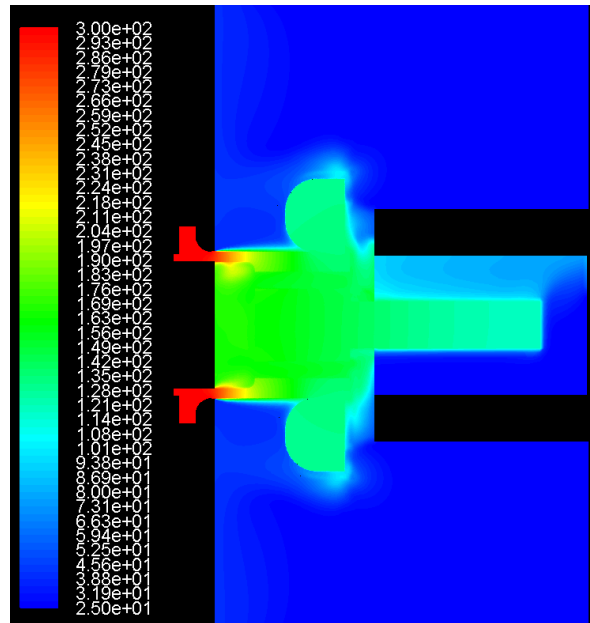


Figure 3. Temperature distribution through cross section of new hub design

In both figures 2 and 3 the boundary condition of 300°C, applied to the section of the models that are in contact with the impeller and located inside the fan casing, is shown in red on the distributions. After the seal point the hub is located outside of the fan case and the temperature distribution through the model reflects this. Immediately after the seal section the colours show a difference between the two models, this is due to the wall thickness of the two geometries, the old hub is 2mm thinner. If a line is taken through both models as shown in figure 4 we can see the temperature distribution through the models.

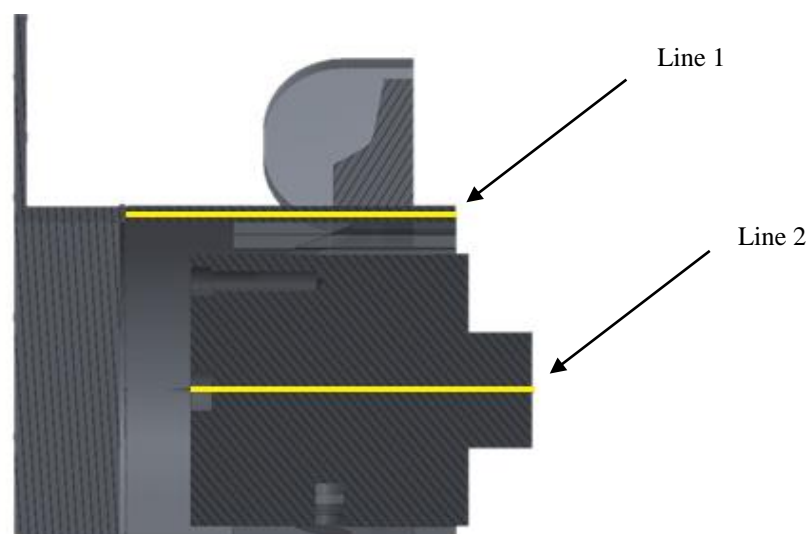


Figure 4. Detail of the two lines used to investigate the temperature distribution through the models

The temperature along the line located within the wall of the hub, (line 1) begins after the seal section in the models, the graph in figure 5 shows the change in temperature through to the front section of the model.

This backs up how the difference in wall thickness is causing the overall temperature to be greater in the new design. The graph details how initially after the seal section, the temperature decreases in a linear trend until a point at which the temperature in the new design does not decrease as quickly as the baseline model. The reason for this lies again in the geometry, if a comparison is made between the models in figure 1 it can be seen that there is a noticeable increase in the material area at the centre of the hub.

Once the baseline model reaches around 120°C and the new model 140°C, there is a definitive levelling out of the temperature on the graph. This point within both models is where the cooling disc ends which suggests that for the cooling disc to be at its most effective it should be located as close to the end of the hub as possible.

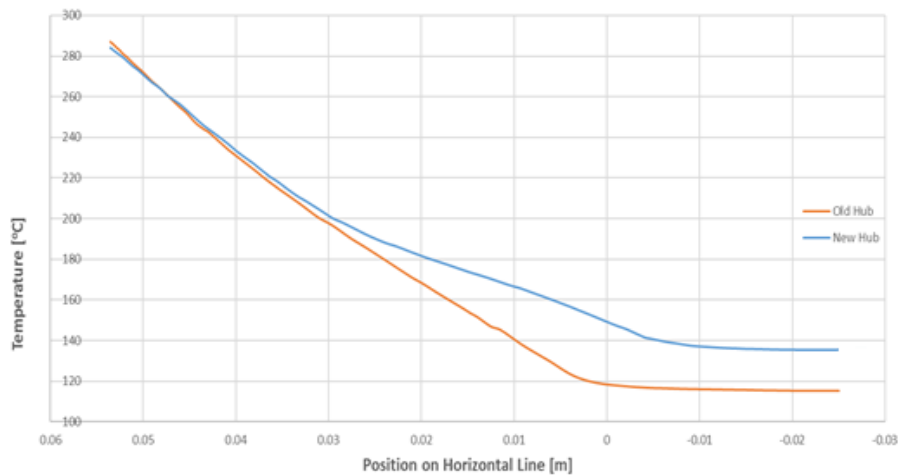


Figure 5. Temperature distribution through line one for both designs

Figure 6 shows the temperature distribution from line 2 in the models, the new model shows a higher overall temperature. This would be due to the thicker geometry sections already discussed which causes a higher temperature at the centre of model. The hub is in use to control the temperature in the shaft so the motor bearing does not become too great, to measure this a point was added where the bearing would sit and is located at the end of line 2. At this point the temperature when using the baseline hub model is 117°C and on the new hub design 135°C, an increase of 13%.

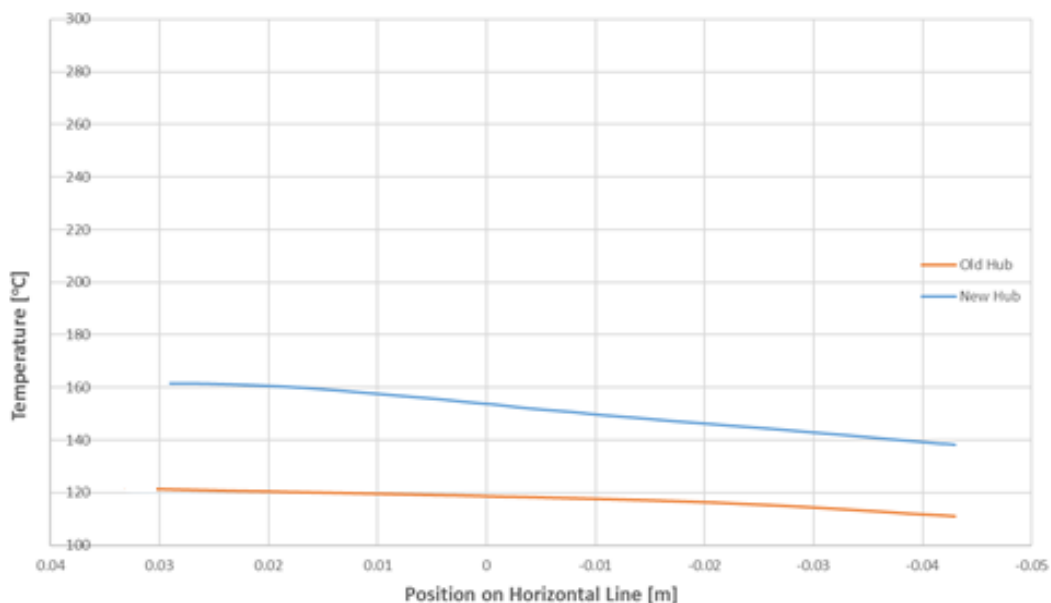


Figure 6. Temperature distribution through line two for both designs

This can pose a potential problem, as the dissipation of heat by new hub is not as effective as the old design, which means that the bearing temperature will increase. The main difference in the geometry is the internal fin arrangement of the baseline design compared to the machined holes of the new design. From the literature reviewed, it was clear to see that existing methods for dissipating heat away from a rotating shaft involve using finned apparatus.

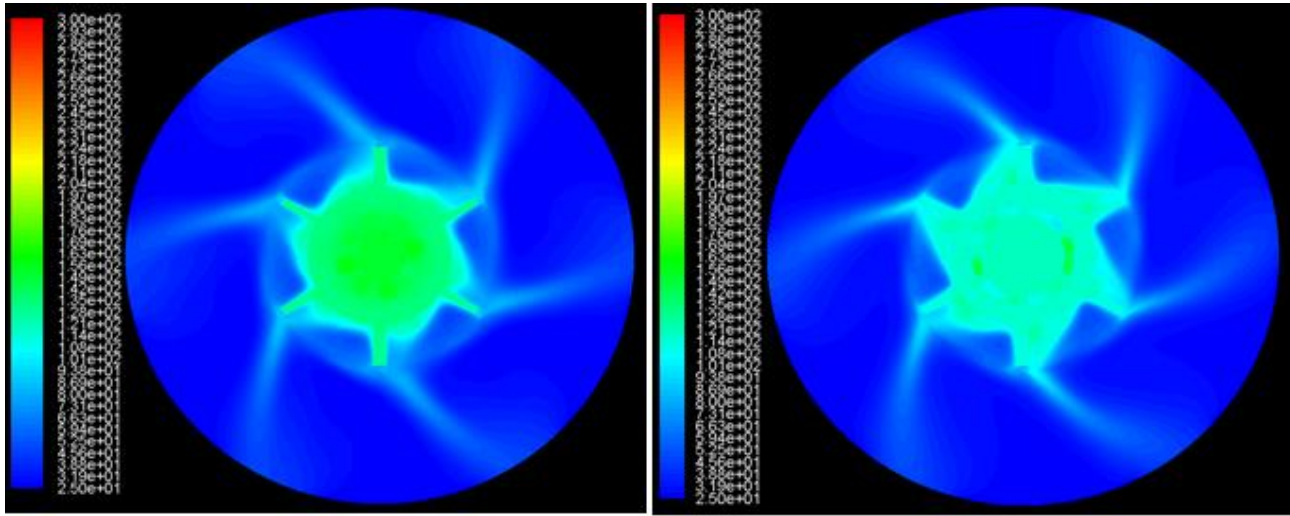


Figure 7. Temperature distribution in the vicinity of the cooling discs

Figure 7 depicts a cross section view through the cooling discs showing the temperature contours, the baseline illustrates a lower overall temperature. However by removing material through machining holes the new design has shown it is possible for other methods to be applied to prevent high temperatures travelling through a shaft, even if the temperature reduction is not as great. This difference is shown in the calculated heat transfer coefficient; this was taken at the edge of the cooling disc for both models. It was found to be $118.18 \text{ W/m}^2\text{k}$ on the baseline model and $31.41 \text{ W/m}^2\text{k}$ on the new hub model. This shows the difference in the amount of heat transfer from the models to the surroundings.

3. CONCLUSION

Whilst the old hub design performed better from the simulation with respect to the temperature at the point where the motor bearing sits on the shaft. The new design can still be used for high temperature applications, something that is preferable due to its cost effective design, the reduction in its heat dissipating capabilities however will need to be factored into the design limitations of the product.

REFERENCES

- [1] L. A. Zainullin, M. V. Kalganov, D. V. Kalganov, V. F. Yarchuk, (2015) Cooling the rotating shaft of a high-temperature furnace fan, *Steel in Translation*, 45 (9) 646-649
- [2] Aziz, A., & Khani, F. (2010). Analytic solutions for a rotating radial fin of rectangular and various convex parabolic profiles. *Communications in Nonlinear Science and Numerical Simulation*, 15(6), 1565-1574.
- [3] Watel, B., Harmand, S., & Desmet, B. (2000). Influence of fin spacing and rotational speed on the convective heat exchanges from a rotating finned tube. *International journal of heat and fluid flow*, 21(2), 221-227.
- [4] Xie, M., Xue, Z., Qu, W., & Li, W. (2015). Experimental investigation of heat transfer performance of rotating heat pipe. *Procedia Engineering*, 99, 746-751.
- [5] Mori, M., Novak, L., & Sekavčnik, M. (2007). Measurements on rotating blades using IR thermography. *Experimental Thermal and Fluid Science*, 32(2), 387-396.

Abrasive Spur Gear Wear Prediction Model: Utilization of the Statistical Design Experiments

Surapol Raadnui & Kitisak Wangwira¹

Department of Production Engineering, Faculty of Engineering
King Mongkut's University of Technology North Bangkok (KMUTNB)
1518 Pracharaj 1 Road, Bang-Sue, Bangkok, Thailand, Postcode 10800

¹ Post Graduate Student, Department of Production Engineering, Faculty of Engineering
King Mongkut's University of Technology North Bangkok (KMUTNB)
1518 Pracharaj 1 Road, Bang-Sue, Bangkok, Thailand, Postcode 10800

ABSTRACT

In extreme spur gear abrasive wear conditions some of the tooth damages such as pitting are not the main type of teeth flank failure any more. For certain abrasive wear limits of teeth flanks, experimental results were presented by statistical parameters. In this particular work the abrasive wear of spur gear lubricated with simulated contaminant was performed. An experiment was carried out which simulates the operation of spur gears in contaminated SiC media. A statistical model and statistical parameters are suitable for the development of abrasive wear model of tested gear pairs. Design of Experiment (DOE) was utilized against the background of spur gear abrasive wear, a multiple correlation model with a limited test conditions was anticipated in this particular paper. The effect of SiC concentration, applied load and sliding distance was statistically and physically analyzed in detail.

Keywords: Design of Experiment; Spur Gear; Three-body Abrasive Wear; Tribology; Wear Modeling Analysis

1. INTRODUCTION

In general, it is well established that the presence of hard contaminants either generated wear particles or ingested outside hard particles in lubricant is inevitable. This undesirable debris can do harmful effect on contacting surfaces and lead to components and system failures. In recent years, several works were undertaken on solid contamination of lubrication. Dwyer Joyce [1] gave an overall picture on significant variations of a wear particle life cycle in the contact. Sari *et. al.* [2] presented the effect of the presence of solid bodies in lubricant during gear surface contact. This specific work showed that the use of a lubricant contaminated by very fine sand particles, leads to significant wear in the first operating period, in zones with a high rate of sliding. Wedevan and Cusano [3] used simulated contact, a smooth ball-plane device where the ball is given a rotational movement on a glass disc, namely; interferometer. They dent the ball surface and then followed the evolution of the dent in the contact and its effect on the oil film thickness. A mathematical model has been developed to predict the abrasive wear behavior of bagasse fiber reinforced polymer composite, AL 6061 and a nonmetallic spur gear [4] to [7]. Against the background of the abrasive wear model literature, for our work though, we are interested in the effect of hard contamination of lubrication on real steel spur gear wear evolution and its effect over the weight loss and surface roughness parameter (R_a) of the worn gear pair according to contaminant concentration, applied torque and the number of cycles of the driving and driven gears.

2. EXPERIMENTS

An experimental test rig was used to perform the rolling-sliding contact experiment. The rig has been designed for the investigation of spur gear pair. The driving gear is run at a rotational speed of 1500 rpm by

an electric motor. The driven gear is actuated in rotation under the driving gear action and loaded by a simple disc braking system. Figure 1 shows the experimental setup schematically. A simple mechanism of disc braking consisting in plate and a braking oil control system is used to apply this torque. Test spur gear pair contact is continuously lubricated by SiC particles mixing with SAE 30 oil circulating system at a rate of 1.60 L/minute. The preparation of the contaminated lubricant is obtained from a mixture of 16 and 32 g/l. of SiC with SAE oil. The continuous agitation of oil is ensured throughout the test. Typical spur gear abrasive wear test rig is shown in figure 2. Tables 1 and 2 show the principal chemical and mechanical characteristics of the test gears. The lubricant used is SAE 30 API CD/SF. It has a kinematic viscosity 100 cSt. at 40°C. It is representative of engine oil, and is often used for industrial gear circulating oil. The used contaminant is SiC. It was cleaned, filtered to mean size of 400 to 500 μm . The chemical and physical characteristic is shown in Table 3. In addition, figure 3 shows an image of SiC particles used in all tests. A full factorial experimental design is used with three design factors of each two levels to describe response of the weight loss, surface roughness parameter and to estimate the parameters in the first order three-body abrasive wear model. Overall $2^3 = 8$ wear experiments, as shown in Table 4, are carried out.

3. RESULTS AND DISCUSSION

The SiC is brittle and has a very high hardness in comparison with a testing gear specimen. Its particles roll or slip into the gearing contact leading to wear and leaving micro-stripes or micro-furrows on the worn gear surfaces. The surfaces wear by “three-body” abrasion in zones, which present a high rate of sliding, and by fatigue, or adhesion surfaces near the pitch diameter. In this last zone, the SiC particles roll and leave furrows or dents on surfaces. For a better understanding of three body abrasion phenomenon in highly contaminated conditions (i.e. with SiC), optical micrographs were taken for the gear tooth. Figures 4 and 5 give, respectively, detailed pictures of surfaces aspect after operation in mild and severe abrasion conditions. The given micrographs, shows abrasive score marks (furrows, stripes and plastic flow) in sliding direction. They are very severe. This wear is caused generally by the existence of SiC particles in lubricant.

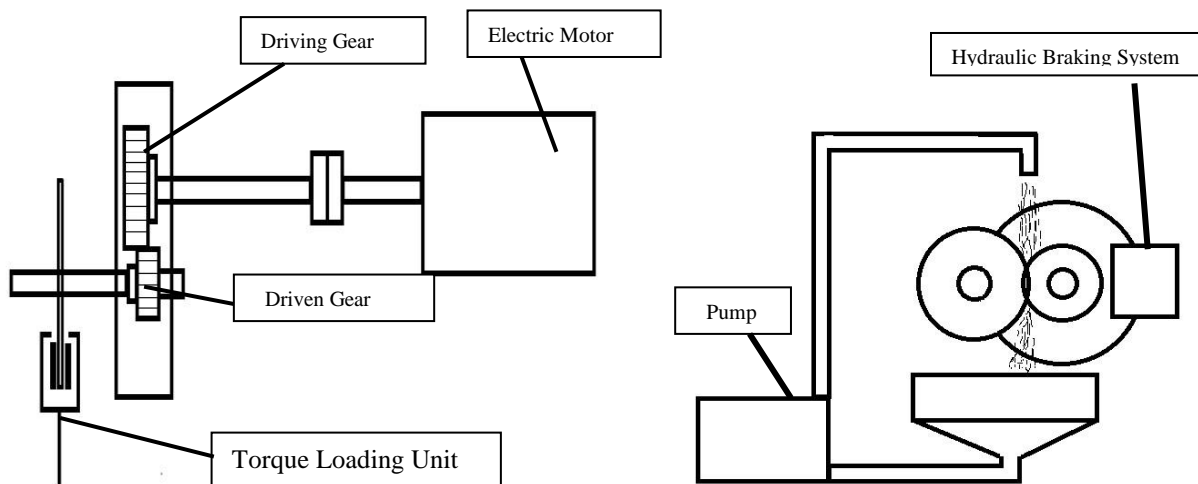


Figure 1. (a) Schematic Diagram of Test Rig. (b) Schematic Diagram of lubricating oil recirculation.

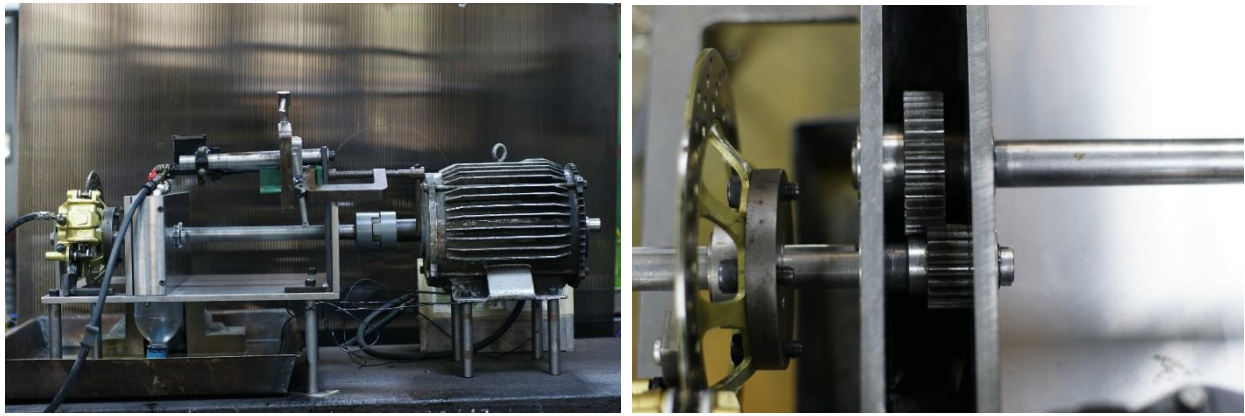


Figure 2. Spur gear abrasive wear test rig.

Table 1: Chemical gear properties

| Gear specimen composition | C | Si | Mn | P | S | Cr | Mo | Ni | Cr+Mo+Ni |
|---------------------------|------|-------|------|--------|--------|------|------|-----|----------|
| | 0.42 | <0.40 | 0.50 | <0.045 | <0.045 | <0.4 | <0.1 | 0.4 | <0.63 |
| | 0.50 | | 0.80 | | | | | | |

Table 2: Mechanical and geometrical gear properties

| Characteristics | Pinion | Gear |
|------------------------------------------------------|--------|------|
| Number of teeth | 35 | 19 |
| Transmission ratio | 1.84 | |
| Height (mm) | 54 | |
| Width (mm) | 20 | |
| Module (2) | 2 | |
| Pitch diameter (mm) | 70 | 38 |
| Pressure angle (α) | 20 | |
| Hardness)HRB(| 87 | |
| Surface roughness parameter: R_a) μm (| 0.773 | |
| Sliding velocity)m/s(| 1.76 | |
| Materials | S45C | |

Table 3: Typical characteristics of SiC media used

| | |
|--------------------------------------|-----------|
| Density (g/cc) | 3.1 |
| Hardness (Kg/mm^2) | 2800 |
| Compressive Strength (MPa) | 3900 |
| Mean Particle Size (μm) | 400 - 500 |

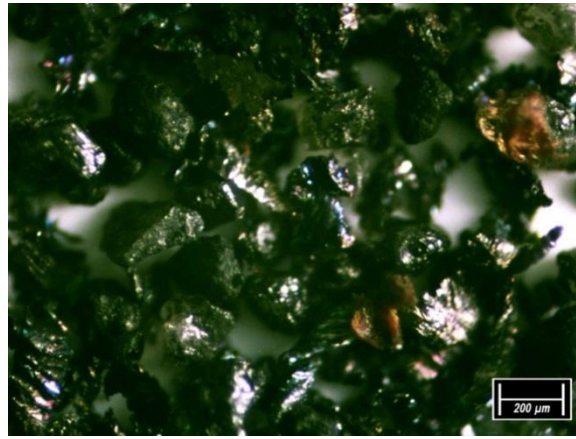


Figure 3. Typical SiC abrasive media used.

Table 4: Factors and their levels for abrasive wear test

| Factor | Level 1 | Level 2 |
|--------------------------|---------|---------|
| SiC concentration (g./l) | 16 | 32 |
| Applied torque (Nm) | 2.98 | 5.37 |
| Sliding distance (m.) | 5000 | 10000 |

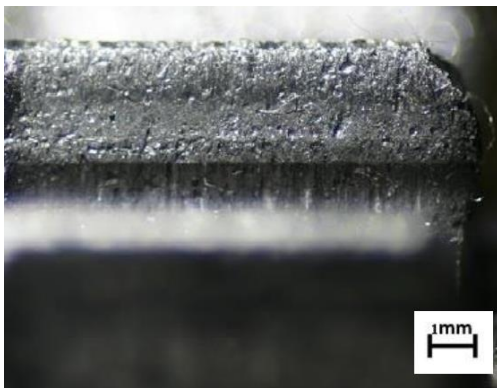


Figure 4. Gear worn surface characteristic: Test No.1.

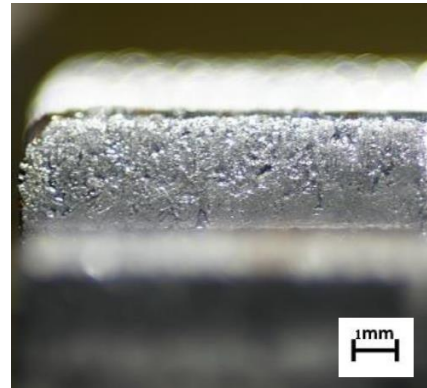


Figure 5. Gear worn surface characteristic: Test No. 8.

By using the full factorial design, a total of 8 experiments are conducted and regression coefficients are calculated. The statistical analysis is made using the popular software used for DOE applications known as MINITAB 16. The full models for three-body abrasive weight loss (g) and surface roughness parameter (R_a = Central Line Average, aka R_{CLA}) can be expressed in term of the coded values of the independent variables, namely; applied torque, SiC concentration and sliding distance in equations (1) and (2) respectively.

$$\text{Weight Loss (mg)} = -2416 + 67.5 \text{ SiC Concentration (g./L)} + 132 \text{ Applied Torque (Nm)} + 0.186 \text{ Sliding Distance (m)} \quad \dots(1)$$

$$\Delta R_a (\mu\text{m}) = 1.45 + 0.0225 \text{ SiC Concentration (g./L)} + 0.114 \text{ Applied Torque (Nm)} + 0.000069 \text{ Sliding Distance (m)} \quad \dots(2)$$

The plot in figure 6 shows that applied torque as the most significant factor for both weight loss and delta R_a . The plots in figure 7 indicate different of interactions between applied torque, SiC concentration and test duration. In general, the amount of interaction is not significant enough to change the importance of the independent factors. This fact finding is also confirmed by the Pareto plots in figure 8. Similar pattern of

results is repeatedly applied for the R_a surface roughness parameter (as another response variable in abrasive wear test), these can be shown in figures 9 to 11.

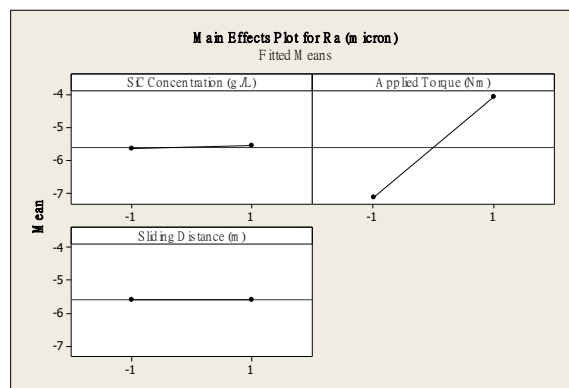


Figure 6. Main effect of abrasive wear factors on weight loss.

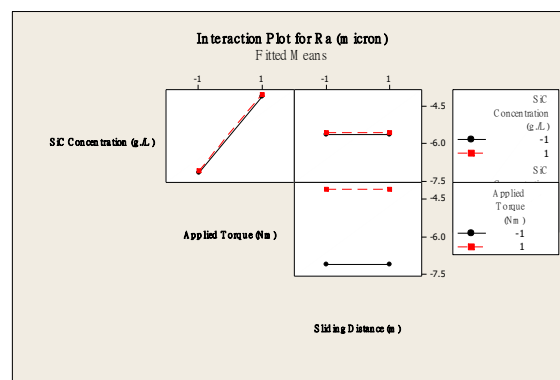


Figure 7. Interaction plot between abrasive wear factors on weight loss.

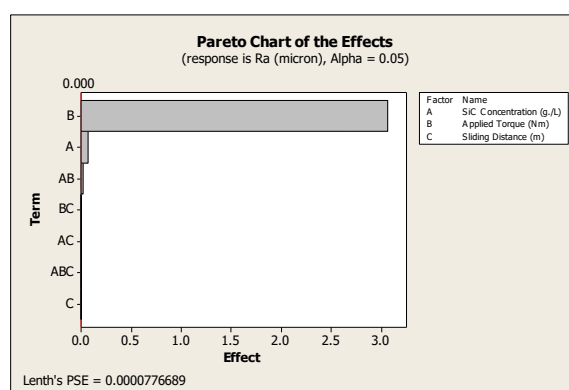
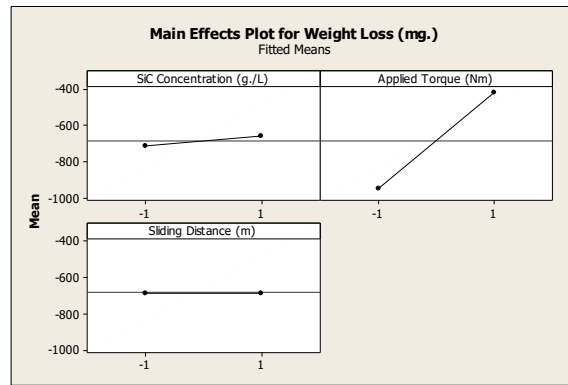
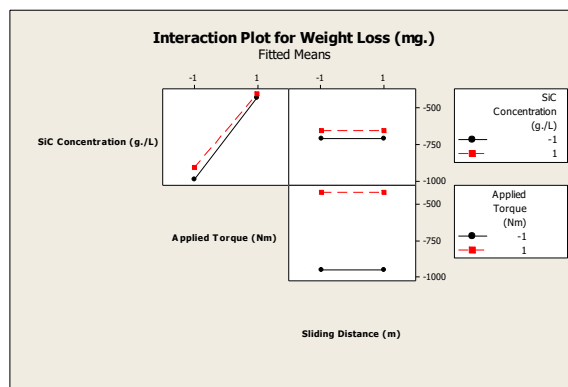
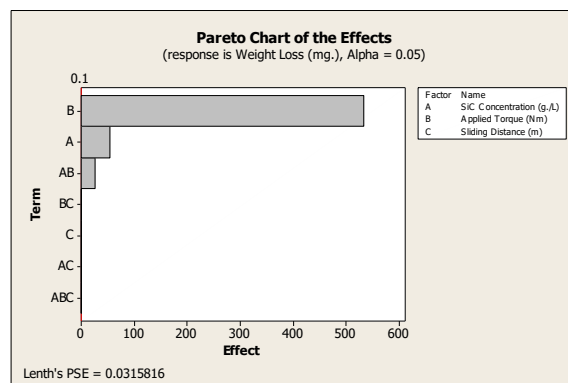


Figure 8. Pareto chart of the main and interaction effects on weight loss.

Figure 9. Main effect of abrasive wear factors on R_a .Figure 10. Interaction plot between abrasive wear factors on R_a .Figure 11. Pareto chart of the main and interaction effects on R_a .

4. CONCLUSION

In this study, full factorial design of experiments has been exploited to develop a first order regression equation for describing three-body abrasive wear behavior of spur gears under SiC contaminated lubricating oil. The relationship of abrasive wear loss and changed surface roughness parameter with applied load, SiC concentration and testing duration has been successfully obtained with a confidence of 95%.

ACKNOWLEDGEMENT

The work reported here is funded by the Thailand Research Fund (TRF) (Grant no. RSA5980067).

REFERENCES

- [1] Dwyer Joyce, R.S. (2004) The life cycle of a debris particle, 31st Leeds-Lyon Symposium on Tribology, Leeds
- [2] Sari, M.R., Haiahem, A. and Flamand, L. (2016) Effect of lubricant contamination on gear wear, *Tribo. Lett.*, 27: 119-126
- [3] Wedeven, L.D. and Cusano, C. (1979) Elastohydrodynamic film thickness measurements of artificially produced surface dents and groove, *ASME J.Tribo.* 22(4) 369-381
- [4] Mishra, P. (2012) Statistical analysis for the abrasive wear behaviour of bagasse fiber reinforced polymer composite, *International Journal of Applied Research in Mechanical Engineering (IJARME)* ISSN: 231-5959, Vol.2, Iss-2
- [5] Khan, M.S., Hasan, Z. and Ansari, Y.A. (2014) Statistical analysis for the abrasive wear behaviour of AL 6061, *Journal of Minerals and Materials Characterization and Engineering*, 2, 292-299
- [6] Sardar, J. and Bandopadhy, (2014) Evaluation of wear behaviour of a non-metallic spur gear, 5th International & 26th All India Manufacturing Technology, Design and Research Conference (AIMTDR 2014) December 12th – 14th, 2014, IIT Guwahati, Assam, India
- [7] Wang, J.C., Chakraborty, J. and Xu, H. (2009) Effects of gear surface parameters on flank wear, *Gear Technology*, 68-73

Degradation-Based Preventive Maintenance Policy for Railway Transport Systems

F. Dinmohammadi^{1*}, B. Alkali¹, M. Shafiee²

¹Department of Mechanical Engineering, School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, United Kingdom, Fateme.Dinmohammadi@gcu.ac.uk

²Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

ABSTRACT

Railway transport maintenance plays an important role in delivering safe, reliable and competitive transport services. It also is one of the major costs for rail transport operations. According to several reports, the inspection and maintenance costs constitute a large portion of the life-cycle cost (LCC) for railway asset infrastructures (such as bridges, rail tracks, track beds and track equipment) and rolling stock components (e.g. chassis, bogies, wheels and wagons). In order to reduce the operating expenditure (OPEX) while maintaining high standards of safety, the asset managers must determine a planning period and find optimum preventive inspection policies for various railway systems, such that the total cost incurred over the life span is minimized and/or the rail network's reliability is maximized. Common railway defects are caused by degradation processes such as rolling contact fatigue (RCF) or wear. The degradation of assets may result in substantial losses to the rail transport operators if it is not prevented in an efficient way. In this paper, we investigate an optimal age-dependent preventive inspection policy for railway assets subject to gradual degradation phenomena. The degradation processes initiate following Non-Homogenous Poisson Process (NHPP) and propagate according to gamma stochastic process. When the size of degradation reaches a critical level, the asset will unexpectedly fail and it has to undergo a corrective repair. This unexpected failure may also interrupt rail operations, cause passenger dissatisfaction or even some accidents like derailment or overturning. To avoid such undesired defects, the asset is preventively inspected at regular time intervals. The problem is to determine an optimal inspection time interval such that the long-run expected cost rate is minimized. The proposed model is applied to support maintenance decision-making for a railway asset on the Scottish rail network. The results show that the use of the proposed inspection policy allows a significant reduction of the maintenance cost compared to the strategy when only corrective repair is considered.

Keywords: Maintenance engineering technologies, Risk assessment, Modelling analysis and optimisation

Corresponding author: Fateme Dinmohammadi (email: Fateme.Dinmohammadi@gcu.ac.uk)

1. INTRODUCTION

The railway transport sector is a key enabler of Britain's economy. The country's rail network with a total length of approximately 16,209 kilometers track is the 18th largest network in the world [1]. The number of rail passengers as well as freight volumes have increased significantly in recent years. According to recent statistics published by the Office of Rail and Road (ORR), a total of 1.654 billion journeys were made in 2014–5, making the UK's railway network the fifth most used in the world [2]. With the growing demand for rail services, the investment on railway has significantly increased during the last decade. Nevertheless, high operation and maintenance (O&M) costs act as a barrier to achieving a favourable financial performance of railway operations. According to several reports, the inspection and maintenance costs constitute a large portion of the life-cycle cost (LCC) for railway asset infrastructures (such as bridges, rail tracks, track beds and track equipment) and rolling stock components (e.g. chassis, bogies, wheels and wagons) [3, 4].

Nowadays, rail transport operators are under increasing pressure to reduce their O&M costs whilst maintaining reliability targets. Railway transport maintenance plays an important role in delivering safe, reliable and competitive transport services as it reduces the potential risk of defects and derailments. Generally, railway defects occur due to a number of specific causes that have been classified by many researchers. Olofsson and Nilsson [5] divided the defects of tracks into two types of surface-initiated and subsurface-initiated cracks. Cannon et al. [6] classified the rail track defects into three main groups: (i) defects originating from rail manufacture, (ii) defects originating from damage caused by inappropriate handling, installation and use, and (iii) defects caused by the exhaustion of the rail steel's inherent resistance to fatigue damage. Dinmohammadi et al. [7] classified the modes of rolling stock defects into six groups, namely electrical faults, structural damages, functional failures, degradation, human errors, and natural (external) hazards. Also, some other classifications have been addressed in reference [8].

The majority of defects in the railway assets are caused by degradation processes such as rolling contact fatigue (RCF), wear, corrosion, erosion, etc. [9]. These degradation processes are very complex as they depend on various factors such as age, traffic density, axle load, asset material, track geometry, curvature, speed, and accumulated Million Gross Tones (MGT) [10]. Any of these forms or their combination can become a cause of a failure.

The degradation of assets may result in substantial losses to the rail transport operators if it is not prevented in an efficient way. Moreover, it may cause accidents [11], traffic disruption and ultimately passenger dissatisfaction. In order to control the rate of rail degradation, age-dependent preventive inspection policies have been extensively used by asset managers [12]. Under this policy, the railway asset is preventively inspected at fixed time intervals kT ($k = 1, 2, \dots$) after its installation. The asset is regarded as failed when the level of its degradation reaches an unacceptable size. In the event of asset failure between two consecutive preventive inspections, a corrective repair has to be undertaken. We assume that the costs for a preventive inspection and a corrective repair tasks are respectively C_0 and C_1 , where $C_1 > C_0 > 0$. The main problem encountered in this policy is to determine the optimal inspection time interval T such that, under given physical/technical constraints, the railroad availability is maximized and/or O&M costs are minimized.

In this paper, we formulate an age-dependent preventive inspection policy for railway assets that are subjected to progressive degradation phenomenon. Degradation processes initiate following Non-Homogenous Poisson Process (NHPP) and propagate according to gamma stochastic process. If the size of degradation reaches a critical level, the asset will unexpectedly fail and it has to undergo a corrective repair. Otherwise, it will be preventively repaired at each inspection epoch. The explicit expression of the long-run expected cost function per unit time for the preventive inspection policy is derived and under certain conditions, the existence and uniqueness of the optimal solution are shown for the infinite-horizon case. The performance of the proposed policy in terms of cost is evaluated and compared to the case when only corrective repair is considered.

The rest of this paper is organized as follows. The assumptions and notation of the model are given in Section 2. In Section 3, we present the problem definition. The formulation of the optimization model and the properties of the optimal solution are discussed in Section 4. In Section 5, the model is applied to a real-life case study. Section 6 concludes this study.

2. ASSUMPTIONS AND NOTATION

In this section, we present the assumptions and notation used in our model formulation.

- The system starts functioning at time zero.
- The decision to repair the railway asset is made based on either its degradation level or its operational age.
- The degradation of the railway asset appears in the form of cracks, wear, corrosion, etc.
- A degradation process initiates following non-homogenous Poisson process (NHPP) and it accumulates gradually.

- The railway asset fails when its degradation size reaches a given threshold D . All defects are assumed to be instantly detected.
- The repair and inspection time is negligible. However, the planned preventive inspection is preferred to an unplanned corrective repair, because the potential passengers can be notified in advance and traffic disruption can be limited.
- The defect threshold D has a pre-specified value, which is determined by the original equipment manufacturer (OEM). The variable T is a decision variable and should be optimized.

| | |
|------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\{N(t): t \geq 0\}$ | number of degradation processes initiating in the interval $[0, t)$ |
| $m(t) [M(t)]$ | intensity [mean value] function of $N(t)$ |
| $P_j(t)$ | the probability that exactly j degradation processes initiate in the interval $[0, t)$ |
| T_j | initiation time of the j^{th} degradation process |
| $F_j(t)$ | cumulative distribution function of T_j |
| $X_j(t)$ | length/size/depth of the j^{th} degradation process t units of time after its initiation |
| D | defect threshold (i.e. critical length/size/depth of degradation) |
| U_j | length of the interval between the initiation time of the j^{th} degradation process to the time that it attains the critical threshold D |
| $g_{U_j}(\cdot)[G_{U_j}(\cdot)]$ | probability density [cumulative distribution] function of U_j |
| $\Gamma(\cdot) [\gamma(\cdot, \cdot)]$ | gamma [incomplete gamma] function |
| S_j | time point that the length/size/depth of the j^{th} degradation exceeds the critical threshold D |
| $S_{[1]}$ | time point that, for the first time, the length/size/depth of a degradation process exceeds the critical threshold D |
| $F_{S_{[1]}}(\cdot)[\bar{F}_{S_{[1]}}(\cdot)]$ | cumulative distribution [survival] function of $S_{[1]}$ |
| $h(\cdot)$ | failure rate function of $S_{[1]}$ |
| $a(x) \bullet b(x)$ | convolution of two functions $a(x)$ and $b(x)$ |
| T | preventive inspection interval |
| C_0 | fixed cost of a preventive inspection |
| C_1 | fixed cost of a corrective repair |
| η | additional cost to the rail infrastructure owner resulting from a corrective replacement |
| $E[X_r]$ | expected length of a maintenance cycle |
| $D(t)$ | s-expected cost of the operating system over $[0, t)$ |
| $CR(T)$ | long-run expected cost per unit time |

3. PROBLEM DEFINITION

Railway assets are subject to various types of degradation modes, such as rolling contact fatigue (RCF), wear, corrosion, erosion, etc. The degradation process for these modes involves three following phases: (i) initiation, (ii) propagation (or growth), and (iii) the failure.

i. Suppose that the degradation processes initiate in the interval $[0, t)$ following a non-homogeneous Poisson process (NHPP), $\{N(t); t \geq 0\}$ with intensity function $m(t)$ and mean value function $M(t)$, i.e.,

$$M(t) = \int_0^t m(x) dx, \quad t \geq 0. \quad (1)$$

where t is the age of the railway asset and $M(t)$ is a non-decreasing function of t with $M(0)=0$. Then, the probability that exactly j ($=0, 1, 2, \dots$) degradation processes occur in the interval $[0, t)$, $P_j(t)$ is given by

$$P_j(t) = P\{N(t) = j\} = e^{-M(t)} \times \frac{[M(t)]^j}{j!}. \quad (2)$$

Let T_j ($j=0, 1, 2, \dots$) denote the initiation time of the j^{th} degradation process in the railway asset, where $T_0 = 0$. We assume that the asset degradation is detected by health monitoring techniques just when they arrive (for more see [13]). Then, the cumulative distribution function of the random variable T_j is given by

$$F_0(t) = 1; F_j(t) = P\{T_j \leq t\} = \sum_{i=j}^{\infty} p_i(t), \quad j = 1, 2, \dots \quad (3)$$

ii. Propagation is the second phase of the degradation process which may be accelerated by adverse environmental conditions. Many models have been developed to study how various degradation processes in different railway assets propagate. For instance, Ringsberg [14] proposed a crack growth model for railway tracks in which the crack propagation life is divided into three stages: (i) shear stress driven initiation at the surface; (ii) transient crack growth behavior; and (iii) subsequent tensile and/or shear driven crack growth (see figure 1).

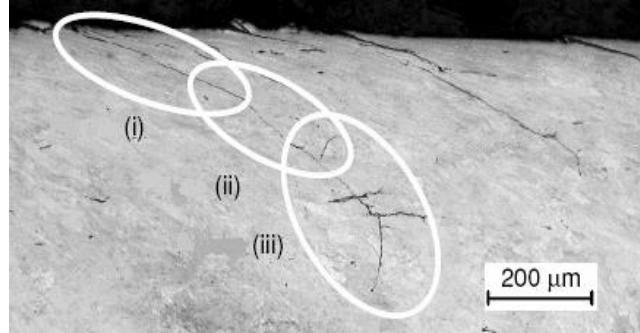


Figure 1. Crack propagation phenomenon in railway tracks.

In this paper, the degradation propagation is modeled using a stochastic *gamma* process, which represents the degradation length/size/depth evolution in time. The gamma process is a stochastic process with independent non-negative increments having a gamma distribution with identical scale parameter. The gamma process has been widely studied for different maintenance applications by several authors (see [15] for a thorough review on the use of gamma process in maintenance modeling). Also, it has been observed that the gamma process is satisfactorily fitted to data of different gradual degradation phenomena (such as wear and crack propagation) in railway industry [16]. Moreover, the existence of an explicit probability distribution function of gamma process permits feasible mathematical developments.

Let $X_j(t)$ be the length/size/depth of the j^{th} degradation process t units of time after its initiation. We assume that $X_j(t)$ has a homogeneous gamma process with shape and scale parameters given by αt and β respectively. Then, for $t > 0$, the density and the cumulative distribution function of the increment of the length/size/depth of the j^{th} degradation process is given by [17]

$$g_{\alpha t, \beta}(x) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-\beta x}, \quad x \geq 0; \quad \alpha, \beta > 0, \quad (4)$$

and

$$G_{\alpha t, \beta}(x) = \frac{\gamma(\alpha t, \beta x)}{\Gamma(\alpha t)}, \quad x \geq 0; \quad \alpha, \beta > 0, \quad (5)$$

where $\Gamma(\cdot)$ [$\gamma(\cdot, \cdot)$] denotes the gamma [incomplete gamma] function, i.e.,

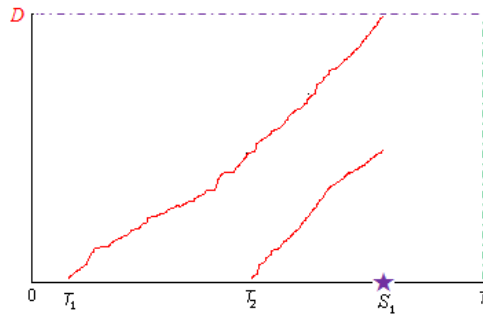
$$\Gamma(v) = \int_0^\infty z^{v-1} e^{-z} dz; \quad \gamma(v, u) = \int_u^\infty z^{v-1} e^{-z} dz, \quad v, u > 0.$$

iii. Railway asset fails when the length/size/depth of a degradation process reaches a given threshold D (see figure 2). In the event of asset failure, a corrective repair is performed and the system returns to an "as-good-as-new condition.

iv.

Let U_j be the length of the interval between the initiation time of the j^{th} degradation process to the time that it attains the critical threshold D , i.e.,

$$U_j = \inf \{t \geq 0 : X_j(t) \geq D\}, \quad j = 1, 2, \dots, \quad (6)$$

Figure 2. Rail asset fails at degradation level D

Then, from Eqs. (4) and (5), the density and the cumulative distribution function of U_j , respectively, are given by

$$g_{U_j}(t) \equiv g_U(t) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} D^{\alpha t-1} e^{-\beta D}, t \geq 0; \quad \alpha, \beta > 0, \quad (7)$$

and

$$G_{U_j}(t) \equiv G_U(t) = \frac{\Gamma(\alpha t, \beta D)}{\Gamma(\alpha t)}, \quad t \geq 0; \quad \alpha, \beta > 0. \quad (8)$$

We denote by S_j the time point that the length/size/depth of the j^{th} degradation process exceeds the critical threshold D . Then,

$$S_j = T_j + U_j, \quad j = 1, 2, \dots \quad (9)$$

Lemma. Let $I_A(\cdot)$ denote the indicator function that is defined as $I_A(\cdot) = 1$ for $x \in A$, and 0 otherwise. Let $\{N_S(t); t \geq 0\}$ be the counting process associated with the random variables S_j ($j = 1, 2, \dots$), that is,

$$N_S(t) = \sum_{j=1}^{\infty} I_{[0,t]}(S_j), \quad (10)$$

Then, having in mind that the convolution of any functions $a(\cdot)$ and $b(\cdot)$ is given by

$$a(x) \bullet b(x) = \int_0^x a(x-t) db(t),$$

$\{N_S(t); t \geq 0\}$ is an NHPP with intensity function,

$$h(t) = m(t) \bullet g_U(t), \quad (11)$$

where $g_U(t)$ is given by Eq. (7) [18].

4. MORE FORMULATION AND ANALYSIS

The railway Asset is preventively inspected and repaired when its operational age attains a value of $T (> 0)$. The cost of a preventive repair is C_0 , whereas the cost of a corrective repair is C_1 . Let $\eta (\geq 0)$ represent the cost parameter referring to an additional cost resulting from an unexpected failure, i.e., $\eta = C_1 - C_0$.

Let X_r denote a maintenance cycle defined by the time interval between maintenance actions (either corrective or preventive). Under the assumptions of the model, we have

$$X_r = \min(S_{[1]}, T), \quad (12)$$

where $S_{[1]}$ denotes the time that, for the first time, a degradation process exceeds the critical threshold D , i.e.,

$$S_{[1]} = \min\{S_j, j = 1, 2, \dots\}, \quad (13)$$

Then, by using lemma, the survival function of $S_{[1]}$ is given by

$$\begin{aligned} \bar{F}_{S_{[1]}}(t) &= P\{S_{[1]} > t\} = P\{N_S(t) = 0\} \\ &= \exp\left(-\int_0^t h(x) dx\right), \end{aligned} \quad (14)$$

where $h(\cdot)$ is the failure rate function of $S_{[1]}$, and is given by Eq. (11). Then, the expected length of a maintenance cycle $E[X_r]$, is given by

$$E[X_r] = \int_0^T \bar{F}_{S_{[1]}}(t) dt, \quad T > 0. \quad (15)$$

Let $D(t)$ be the s -expected cost of operating the system for the time interval $[0, t)$. From the *renewal reward theorem* (see [19, p. 52]), the expected cost rate, denoted by $CR(t)$, is the expected operational cost incurred in a maintenance cycle divided by the expected cycle length, i.e.,

$$CR(T) = \lim_{t \rightarrow \infty} \frac{D(t)}{t} = \frac{(C_0 + \eta) \times F_{S_{[1]}}(T) + C_T \times \bar{F}_{S_{[1]}}(T)}{\int_0^T \bar{F}_{S_{[1]}}(t) dt}, \quad (16)$$

where $F_{S_{[1]}}(\cdot)$ [$\bar{F}_{S_{[1]}}(\cdot)$] is the cumulative distribution [survival] function of $S_{[1]}$. The problem is to find the optimal value of T^* that minimizes the objective function $CR(T)$, given in Eq. (16). Therefore, the proposed optimization model can be formulated as follows:

$$\text{minimise } CR(T) = \frac{C_0 + \int_0^T \eta h(t) \bar{F}_{S_{[1]}}(t) dt}{\int_0^T \bar{F}_{S_{[1]}}(t) dt}, \quad 0 < T \leq T_{\max}, \quad (17)$$

The constraint $T \leq T_{\max}$ implies that for safety requirements or due to the presence of physical/technical constraints (such as technology obsolescence and design modifications), the inspection time interval should not exceed some finite upper limit. The following theorem solves this problem.

Theorem. If $h(T)$ is strictly increasing in t , and $\eta h(T_{\max}) > CR(T_{\max})$, there exists a unique and finite minimum $T^* \in (0, T_{\max})$ that verifies the following equation:

$$\bar{F}_{S_{[1]}}(T^*) + h(T^*) \times \int_0^{T^*} \bar{F}_{S_{[1]}}(t) dt = \frac{C_0}{\eta} + 1, \quad (18)$$

whereas, if $h(T)$ is non-decreasing in t , and $\eta h(T_{\max}) \leq CR(T_{\max})$, then $T^* = T_{\max}$ (implying maximum preventive replacement interval).

Proof. The single-variable optimization model in Eq. (17) is a special case of the framework studied by Aven [20, pp. 151–152]. The optimal T^* can be obtained by differentiating $CR(T)$ with respect to T and setting it equal to zero, if it is an interior point of the feasible region. If none of the solutions of Eq. (17) is within the feasible region, we need to investigate the behavior of $CR(T)$ over the feasible region. If $CR(T)$ is a decreasing function of T , then the optimal preventive inspection interval should be set to T_{\max} .

Remark. Suppose that the degradation processes initiate following a homogeneous Poisson process with constant rate m (>0). Then, from Eq. (11), we have

$$h(T) = mG_U(T). \quad (19)$$

Also, suppose that there is neither physical nor technical constraints on the preventive inspection interval time, i.e., $T_{\max} \rightarrow \infty$. Now, if $\int_0^\infty \bar{F}_{S_{[1]}}(t) dt > \frac{C_0 + \eta}{m\eta}$, there exists a unique and finite minimum optimal age T^* that minimizes the function $CR(T)$, whereas, if $\int_0^\infty \bar{F}_{S_{[1]}}(t) dt \leq \frac{C_0 + \eta}{m\eta}$, the optimal maintenance policy will be repairing the failed railway asset at its critical degradation level D .

5. A CASE STUDY

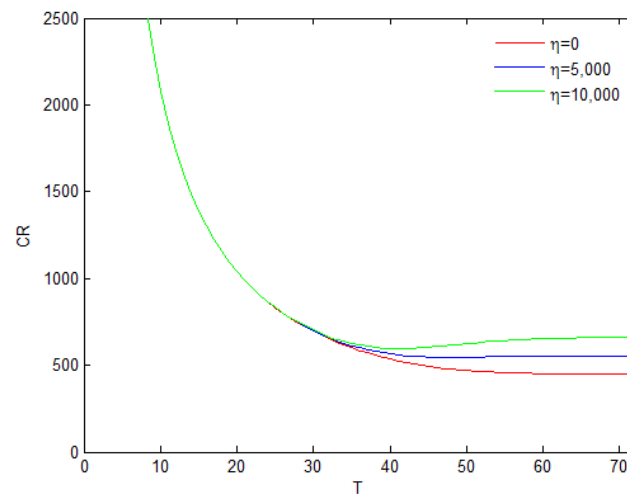
In this section, we present an application of the proposed inspection policy to the Europe's only heavy haul line, Iron Ore Line (Malmbanan). The Malmbanan is a 473km/294mi railway freight line in northern Sweden that runs from Luleå via Gällivare and Kiruna to Narvik in Norway (figure 3).



Figure 3. Iron Ore Line (Malmbanan)

The data has been collected from the literature [10, 21-25]. We assume that the arrival of cracks to the rail track follows a Poisson process with rate $\hat{m} = 0.144/\text{month}$ (i.e., the mean-time-to-initiate a crack is around seven months). The length of the cracks follows a gamma process with parameters $\hat{\alpha} = 0.576$ and $\hat{\beta} = 1.50$ (α/β is 0.384mm per month). The railway track breaks when the length of the crack exceeds the rail web thickness, i.e., $D=16.5\text{mm}$.

Average length of the rail replacement after a break is $L = 8$ meters. The cost of 60E1 railway track (including neutralization) per meter is 2,250 SEK. Average labour cost per hour (including the track worker cost, track welder cost, and inspection personnel cost) is 625 SEK. The hourly rate of hiring the welding equipment or service vessels for maintenance, replace or inspection of the railway track is 80 SEK. The mean time required to perform a maintenance (either corrective or preventive) is 4 hours. However, the corrective type may cause traffic disruption that incurs an additional cost η to the route operator. The physical lifetime of 60E1 railway track is considered to be equal to six years (72 months). We wrote a MATLAB program for the minimization of the expected cost rate, as given in Eq. (17). The pictorial representation of the expected cost rate as a function of the inspection interval T for three different values of $\eta = 0$, $\eta = 5,000$, and $\eta = 10,000$ SEK is shown in figure 4.

Figure 4. Expected cost rate for different values of η .

From figure 4, it is found that the use of optimal age-dependent preventive inspection policy allows a significant reduction of the maintenance cost compared to the strategy when only corrective repair is considered (the corresponding cost is the asymptote of the path, when T tends to T_{max}). The percentage reduction of the maintenance cost achieved through applying the optimal age-dependent preventive inspection policy is obtained as

$$r = \frac{CR(T_{max}) - CR(T^*)}{CR(T_{max})} \times 100. \quad (20)$$

The optimal value of T^* and the corresponding expected cost rate, $CR(T^*)$, the expected cost rate for corrective maintenance policy, $CR(T_{max})$, and the percentage reduction of the maintenance cost, r are presented in Table 1. It can be seen that as the cost parameter η increases, the optimal inspection interval T^* becomes shorter, however the expected cost rate, $CR(T^*)$ increases. Also, when a large additional cost is likely to be incurred by the infrastructure owner in corrective maintenance case, applying the age-dependent preventive inspection policy will be more efficient than corrective repair and has a huge potential to reduce the maintenance cost. For instance, when the cost parameter η is 10,000, the age-dependent policy allows for approximately %10 reduction of the maintenance cost compared to the corrective repair policy.

Table 1: Results of the optimization model for different values of η .

| | Unit | $\eta=0$ | $\eta=5,000$ | $\eta=10,000$ |
|---------------|-----------|----------|--------------|---------------|
| T^* | month | 72 | 49 | 42 |
| $CR(T^*)$ | SEK/month | 448.38 | 546.40 | 594.81 |
| $CR(T_{max})$ | SEK/month | 448.38 | 554.94 | 661.49 |
| r | % | 0 | 1.54 | 10.08 |

6. CONCLUSION

In this paper, an optimal age-dependent preventive inspection policy is presented for railway assets subject to progressive degradation. Under this policy, the railway asset is either preventively repaired at fixed time intervals kT ($k = 1, 2, \dots$) or it undergoes a corrective repair action at an unacceptable degradation level D . This study can be extended in many directions to make it more practical in maintenance management of railway industry. Some of the possible extensions are:

- In this paper, we assumed that the cost discount rate (the time value of money) is zero. More work is needed to investigate the optimal solution for a discounted case with a positive cost discount rate α ;
- Formulating and analyzing the model when the degradation can be only detected if their length/size/depth reaches a detection threshold c (>0); and finally,
- Providing a cost comparison of the proposed age-dependent inspection policy with other common strategies such as reliability-centred maintenance (RCM).

We have worked on some of these extensions and our findings will be reported in the near future.

REFERENCES

- [1] <http://www.telegraph.co.uk>, accessed on June 18th, 2016.
- [2] Office of Rail and Road (ORR) (2015) Passenger rail usage 2014-15, Quarter 4. Available online: <http://orr.gov.uk>.
- [3] Patra, A.P. (2009) *Maintenance decision support models for railway infrastructure using RAMS & LCC analyses*. Doctoral Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.
- [4] Dinmohammadi, F., Alkali, B., Shafiee, M., Berenguer, C. and Labib, A. (2016) Risk evaluation of railway rolling stock failures using FMECA technique: a case study of passenger door system. *Urban Rail Transit* 2(3), 128–145.
- [5] Olofsson, U. and Nilsson, R. (2002). Surface cracks and wear of rail: a full-scale test on a commuter train track, *In Proc. of the Institution of Mechanical Engineers*, 216(4), 249–264.
- [6] Cannon, D.F., Edel, K.O., Grassie, S. L. and Sawley, K. (2003) Rail defects: an overview, *Fatigue & Fracture of Engineering Materials & Structures* 26(10), 865–886.
- [7] Dinmohammadi, F., Alkali, B., Shafiee, M. (2016) A risk-based model for inspection and maintenance of railway rolling stock. In: *European Safety and Reliability Conference (ESREL)*, Glasgow, UK, 25.9.2016– 29.9.2016.
- [8] Kumar, S. (2006) *A study of the rail degradation process to predict rail breaks*. Licentiate Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.
- [9] Kumar, S. (2008) *Reliability analysis and cost modeling of degrading systems*. Doctoral Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.

- [10] Shafiee, M., Patriksson, M., Chukova, S. (2016) An optimal age-usage maintenance strategy containing a failure penalty for application to railway tracks. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 230(2), 407–417.
- [11] Holmgren, M. (2005) Maintenance-related losses at the Swedish Rail. *Journal of Quality in Maintenance Engineering* 11, 5–18.
- [12] Shafiee, M. and Finkelstein, M. (2015) An optimal age-based group maintenance policy for multi-unit degrading systems. *Reliability Engineering & System Safety* 134, 230–238.
- [13] Hopkins, B.M. and Taheri, S. (2010) Track health monitoring wavelets, in *Proceedings of the ASME Rail Transportation Division Fall Technical Conference, October 12–13, Roanoke, Virginia, USA*.
- [14] Ringsberg, J.W. (2001) Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue* 23, 575–586.
- [15] van Noortwijk, J.M. (2009) A survey of the application of gamma processes in maintenance. *Reliability Engineering and System Safety* 94, 2–21.
- [16] Meier-Hirmer, C., Sourget, F. and Roussignol, M. (2005) Optimising the strategy of track maintenance. In *Proceedings of the European Safety and Reliability Conference (ESREL), June 27–30, Tri City, Poland*, pp. 1385–91.
- [17] Park, C. and Padgett, W.J. (2008) Cumulative damage models based on gamma processes. *Encyclopedia of Statistics in Quality and Reliability*. DOI: 10.1002/9780470061572.eqr119.
- [18] Shafiee, M. and Finkelstein, M. (2015) A proactive group maintenance policy for continuously monitored deteriorating systems: application to offshore wind turbines. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 229(5), 373–384.
- [19] Ross, S.M., *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, CA, 1970.
- [20] Aven, T. (1992). *Reliability and risk analysis*. Elsevier Applied Science, London.
- [21] Schöch, W. and Frick, A. (2012) Optimizing rail grinding strategies on the Malmaban. *Railway Gazette International* 168(1), p. 46.
- [22] Patra, A.P., Söderholm, P. and Kumar, U. (2009) Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration. *Institution of Mechanical Engineers. Proceedings. Part F: Journal of Rail and Rapid Transit*, 223(3), 285–293.
- [23] Kumar, S. (2008) *Reliability analysis and cost modeling of degrading systems*. Doctoral Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.
- [24] Nordmark, T. and Larsson-Kraik, P-O. (2007) Managing rail degradation on the Malmaban. *Railway Gazette International* 163(6), p. 46.
- [25] Reddy, V., Chattopadhyay, G., Larsson-Kraik, P-O. and Hargreaves, D.J. (2007) Modelling and analysis of rail maintenance cost. *International Journal of Production Economics* 105, 475–482.

Condition-Based Maintenance Decision Making: a Practical Approach for Marine Vessels

Lilach Goren Huber¹, Simon Kunz² and Marcel Dettling¹

¹ Zurich University of Applied Sciences, Winterthur, Switzerland.

² MESPAS AG, Zurich, Switzerland

ABSTRACT

Condition monitoring and predictive maintenance are gaining importance amongst researchers and practitioners. Many technical systems, however, do not get to profit from the growing amount of knowledge in these fields. The reason lies in the lack of high quality monitoring data, largely due to the reluctance of system owners to invest in monitoring equipment and data processing. This leads to a vicious circle that limits the applicability of condition based maintenance. The need for simple yet effective schemes for data-driven maintenance optimization motivated the present research. This paper demonstrates the development of such a scheme for maintenance optimization of marine vessels based on the readily available data, thereby lowering the entry barrier for ship operators and promoting future improvements of the data quality.

The approach combines data-driven with physical modeling, and uses the models as an input for a constrained optimization algorithm based on linear programming and a Monte-Carlo-based uncertainty analysis. The algorithm allows for a gradual enhancement of the data-driven component upon acquisition of more and better data.

Keywords: Condition Based Maintenance, Decision Making, Modelling and Optimization, Degradation Modelling, Uncertainty analysis.

Corresponding author: L. Goren Huber (email: lilach.gorenhuber@zhaw.ch)

1. INTRODUCTION

The main goal of Condition Based Maintenance (CBM) is to assess the equipment condition in order to make appropriate maintenance decisions and reduce maintenance and related costs [1]. Future-Condition-Prediction-Based method (FCPB) is a method of CBM decision-making, aimed at predicting future equipment condition and planning appropriate maintenance accordingly. This framework consists of several steps [2]: i) equipment deterioration modelling ii) future equipment condition forecasting and iii) maintenance activities scheduling.

The number of FCPB applications has been growing in the past years. Most of them are centered at deterioration modelling towards predicting future equipment condition. Only few of them deal with the issue of integrating the modelling in a decision-making process [2]. The latter aspect is however crucial for practitioners if they are to effectively use the information about the equipment condition. A central challenge in this aspect is to construct models that can gain acceptance amongst practitioners.

The main body of research in the field of FCPB (and CBM in general) focuses on equipment with large amounts of monitored data. Most case studies in the literature deal with engine systems [3], pumps, gear systems [4], bearings [5], compressors [6] and turbines [7]. Applications of CBM methods to other systems are limited, mainly due to lack of reliable monitoring data. This, however, poses a difficulty to encourage practitioners to collect condition monitoring data, as the acquaintance and trust in CBM methods is limited. Thus, many application areas where condition monitoring and CBM would actually be of great financial

benefit, neglect this option altogether. This increases the gap between research and practice in these areas even further.

In this paper we suggest a general approach that is aimed at breaking this “vicious circle”: one starts out with a relatively limited amount (and possibly quality) of data out of already existing basic monitoring systems. Based on these data one develops a degradation model and a decision-making framework which are largely supported by prior knowledge collected and documented in the domain specific literature. The integration of the condition monitoring data and the literature information is done with the help of Bayesian statistical methods [8]. Such a decision-making framework has to meet at least three requirements in order to gain practitioners acceptance: i) simplicity of modelling, also enforced by the low data quantity/quality ii) robustness of the model against small modifications in the input parameters iii) mechanism for assessment of uncertainty and sensitivity analysis which are transparent to the user. Once these criteria are fulfilled, the chances are higher that the method will be accepted and will lead to a stronger motivation for improved data acquisition in the future. The newly collected data will be fed back into the model, improving its reliability and reducing uncertainty. This iterative approach to implementing CBM decision-making methods for industrial applications might open the door for an increasing number of application areas into the world of data-driven maintenance.

In this paper we demonstrate the approach for the CBM decision-making for hull and propeller of large marine vessels (e.g. cargo ships, oil tankers, etc.). We focus on a deterioration process which comes about due to the attachment of marine organisms to the ship hull and propeller, thereby increasing its resistance in water, the fuel cost and the emission of greenhouse gases. This system differs from many industrial equipment systems in the fact that the main effect of the condition degradation is an increase in operating costs rather than leading to a failed state that requires reactive maintenance measures. Hence, predicting the time to failure is not a relevant topic here, but rather predicting the future degradation state of the system. We believe that the problem of gradually increasing energy consumption due to a deteriorated condition can be relevant to cost-effective maintenance planning in many other systems and is therefore of broader interest.

The approach covers both steps of CBM decision making, including deterioration modelling for the prediction of the future equipment condition and optimization of decision-making for the planned maintenance in order to minimize the sum of operation and maintenance (O&M) costs over a finite planning horizon. In addition the approach includes a mechanism for uncertainty analysis for the decision-making results. The latter is a step with important implications for the acceptance of practitioners and yet is rarely integrated in the commonly reported approaches for CBM decision-making [9-11].

The paper is divided into the following sections: 1. Introduction to the case study 2. Deterioration modelling approach 3. CBM decision-making: the optimization framework 4. Framework for uncertainty analysis.

2. INTRODUCTION TO THE CASE STUDY: CBM OF SHIP HULL AND PROPELLER

The operation of large ocean going vessels requires an enormous amount of energy in the form of fuel oil. This entails the largest cost factor for vessel operations, with over 20M\$ annually for a containership at current oil prices. In attempts to reduce operation expenses, the shipping industry is making an effort to address the prominent reasons for increased fuel consumption [12]. One such reason is the accumulation of marine organisms on the wetted surface of the hull and the propeller of a vessel, also known as fouling. Fouling on a ship's hull significantly increases drag, reducing the overall performance of the vessel, and increasing the fuel consumption, when trying to maintain speed and schedule [13]. According to investigations done on various vessel types, the added resistance due to fouling can grow up to 40% over a year [14].

The high fuel costs and increased environmental regulations are forcing operators of all vessel types to search for ways to monitor fouling conditions and cost-effectively counteract its effect [12]. A central challenge in trying to do so is the difficulty to measure and quantify fouling. Estimating the fouling

condition of hull and propeller and predicting their deterioration over time can only be done indirectly through monitoring and quantification of the added fuel consumption.

A further challenge is to use the fouling condition monitoring and prediction in order to plan the maintenance of the ship hull and propeller in a cost-effective way. Once the fouling level increases there are several measures to reduce it: (1) under water propeller polishing (2) under water hull and propeller cleaning (3) hull polishing and anti-fouling re-painting at dry dock. The three maintenance actions are ordered according to their cost factor, but also according to their impact on the fouling condition of the vessel. There is therefore a trade-off between the saved operational costs for fuel and the maintenance and downtime expenses in choosing between different maintenance actions.

In principle, any one of the three maintenance actions could be scheduled at any time during the planning horizon (which realistically spans around 1-2 years). In practice, some constraints are imposed by the needs and requirements of the ship operators. Some examples include prescheduled maintenance actions, time windows of unavailability of the vessel for maintenance service, and minimal time between actions. The goal of the decision making approach we present is to select the most cost-effective maintenance plan which is achievable under all given constraints.

The optimization framework for maintenance scheduling requires inputs from the deterioration model: the rate of fouling condition degradation and its impact on the operational cost, and the effect of the three maintenance actions in reducing the fouling. The formulation of an adequate deterioration model is necessary for the implementation of a maintenance optimization framework which is applicable under real-world conditions. In the next section we describe the deterioration model and the assumptions behind it.

3. DETERIORATION MODEL

The goal of the deterioration model is to estimate the time dependent added fuel consumption as a result of hull and propeller fouling, and to predict the future evolution of this added fuel consumption. To this end, we combine physical and data-driven modelling approaches: we construct a regression model, where the choice of regression variables is supported by physical understanding of ship propulsion and resistance.

Nomenclature

| | |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------|
| $y_{fuel}, x_{speed}, x_{draft}$ | Fuel consumption, speed and draft |
| $x_{whead}, x_{wside}, x_{wtail}$ | Wind speed in head/side/tail direction as in Ref. [15] |
| x_{pp}, x_{hc}, x_{dd} | Fouling regression variables for propeller polishing, hull cleaning and dry dock |
| $\beta_1, \beta_2, \beta_3$ | Slopes of fouling contributions to the added fuel consumption (for <i>pp</i> , <i>hc</i> , <i>dd</i> respectively) |
| Δt | Time resolution of the discretized problem (here one week) |
| t_H | Maintenance planning horizon (in weeks) |
| I | Set of time periods 1... t_H (weeks) |
| D | Set of maintenance actions 1,2,3 (<i>pp</i> , <i>hc</i> , <i>dd</i> respectively) |
| C_{Fuel} | Fuel price |
| C_{M1}, C_{M2}, C_{M3} | Costs of maintenance actions 1,2,3 |
| t_{na} | Set of time periods of unavailability for maintenance |
| t_{min} | Minimal time between maintenance actions (in weeks) |
| A_f, T_f | Set of actions and time periods which are prescheduled by the user |

3.1. Condition monitoring data availability

As explained above, we adopt an approach of minimal effort for new data acquisition in order to enhance acceptance of the framework by practitioners. The regression model is therefore based on data which are

already available for ship operators in the form of “noon reports” of the ship crew on board, including daily fuel consumption, mean draft and vessel speed. This is supported by satellite position data of the vessel which can cheaply be acquired from commercial service providers. They enhance the accuracy of the speed reading and allow integration of weather conditions and sea state (such as wind speed and direction) out of readily available low-cost services. The data are typically collected over a period of several years in order to enable the detection of the temporal evolution of fouling.

3.2. Data selection and regression model

Our deterioration model starts with a pre-selection step. We restrict to days where the vessel was continuously cruising in open sea and also avoid situations with adverse weather conditions. This process usually retains around 30% of observations for large ocean-going vessels and has the benefit that we do not need to model complicated processes such as acceleration, maneuvering, etc. This allows us to highly simplify the regression model and to include a minimal set of covariates only, whose role is relatively well understood by the domain specific literature [16]. Because we are interested in the long-term evolution of fouling, the benefits of the pre-selection step by far outweigh the associated loss of data.

Our regression model relates the daily fuel consumption to aggregated daily values for speed, draft, wind and the fouling condition. Other contributing factors such as sea temperature, water salinity, waves and sea depth certainly play their role but were found to be largely irrelevant on our preselected data. The proportion of explained variance in the fuel consumption by our models usually ranged around 90-95%. For operationalizing the regression model, some further engineering and feature construction were required. These were guided by physical understanding of fouling and constraints in data accessibility. In particular:

- We construct a deterministic rather than a stochastic deterioration model. Due to data scarcity we do not aim at capturing the stochastic properties of the deterioration process but only the overall degradation trend over time.
- For practical reasons, we model the logarithm of the fuel consumption. This allows capturing and efficiently estimating the associated power laws for its dependence on speed, draft and wind. Furthermore, fouling becomes a multiplicative penalty to fuel consumption (i.e. creates e.g. 10% of added fuel consumption at arbitrary speed). A purely additive model would not work.
- We assume that the fouling contributions grow linearly with time on the scale of the logged fuel consumption. They can grow indefinitely as long as no maintenance actions are performed. These assumptions create a strong simplification, but are well backed-up by the literature ([14], [17]) for short-term analysis of vessels that are in service and under repeated maintenance.
- We assume fouling contributions of hull and propeller which can be eliminated by hull and propeller cleaning respectively. In addition, there is a third fouling component only to be removed by dry dock maintenance which brings the vessel to “as good as new” condition with respect to fuel consumption. The other two actions have an imperfect effect which reflects current field knowledge (cf. [14]).
- The effect for each of the fouling contributions and the respective maintenance actions are estimated from the data. However, we assume that both the severity of the temporal degradation and the benefit of maintenance do not change over time. This assumption is based on evidence in the literature under the assumption of careful but effective cleaning with no damage to the anti-fouling coating of the hull, see [14], [17]. This assumption also yields the basis for predicting the future evolution of fouling by extrapolation of the estimated model.

It is important to notice that the temporal degradation due to fouling can finally be expressed by 3 parameters only $\beta_1, \beta_2, \beta_3$ which are the essential input for the optimization algorithm.

The regression model we fit to the data is therefore

$$\log(y_{fuel}) \sim \log(x_{speed}) + \log(x_{draft}) + x_{whead} + x_{wside} + x_{wtail} + x_{pp} + x_{hc} + x_{dd} \quad (1)$$

This model is fitted using the OLS algorithm under assumption of independent Gaussian errors. Residual analysis shows that these assumptions are justified and the practical difference by employing more robust or more general procedures was negligible. An example of the regression results is shown in figure 1. It shows partial residuals versus time, i.e. the evolution of the fuel consumption with the estimated effects of speed, draft and wind removed. What remains besides the unavoidable scatter is the contribution of fouling and the effect of maintenance. The solid red line reports the expected fuel consumption for which a bootstrap-based confidence region (highlighted in grey) is provided. Hence, we identified an approach for estimating the current and future fouling conditions of any vessel which can serve as an input for maintenance optimization. The next section explains how we combine these results with prior knowledge from the domain specific literature in order to further robustify our approach.

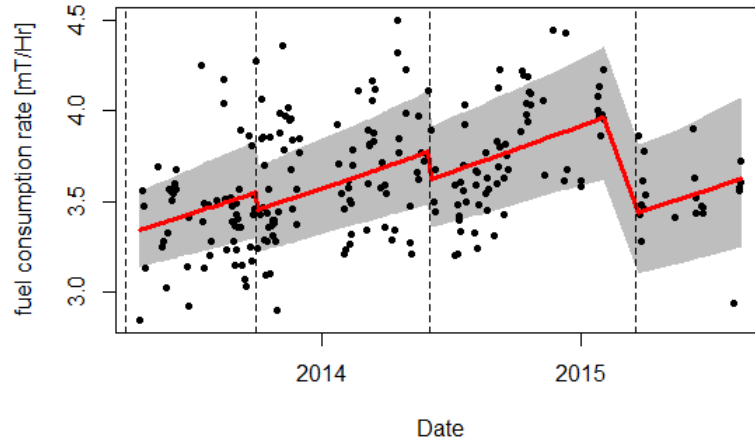


Figure 1. Deterioration model, results of regression analysis

3.3. PDF of the model parameters: Bayesian approach

In order to improve the reliability of the results of the degradation model, we combine prior knowledge collected from previous observations on hull and propeller fouling effects and maintenance [14], [18]. Out of this information we derive an approximated prior $P(w)$ for the three fouling-related parameters of the regression model. This happens under the assumption that the variables are all independent and normally distributed, $P(w) \sim N(\mu_0, S_0)$ with the mean μ_0 and covariance S_0 derived from the literature. Along with the prior distribution we use the observed response variable y , the design matrix X with the Gaussian noise error $\varepsilon \sim N(0, \sigma^2)$. The calculation of the posterior distribution $P(w|y) \propto P(y|w)P(w)$ follows the guidelines in [19], and yields $P(w|y) \sim N(\mu, S)$ with

$$\begin{aligned} S^{-1} &= S_0^{-1} + \frac{1}{\sigma^2} X^T X \\ \mu &= S \left(S_0^{-1} \mu_0 + \frac{1}{\sigma^2} X^T y \right) \end{aligned} \quad (2)$$

Note that in our case, we use priors for the fouling coefficients only. In this case we derive the reduced covariance matrix by integrating out all degrees of freedom which are not related to fouling in the original covariance matrix.

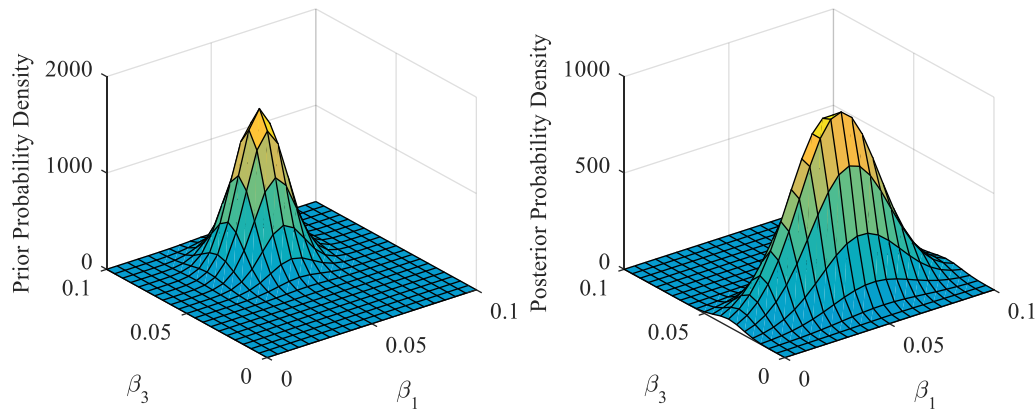


Figure 2. Prior and posterior distribution functions for the degradation model parameters β_1 and β_3

An example for the prior and posterior distributions is shown in figure 2. In this case the data provided information only about the model parameters β_1 and β_3 , and the distribution of β_2 is derived from prior knowledge only. The mean values of the posterior μ will be used by the optimization algorithm to calculate an optimal maintenance policy. The posterior distribution $P(w|y)$ will be used for an uncertainty analysis of the optimization outcomes.

4. MAINTENANCE OPTIMIZATION

4.1. Problem definition

The degradation model suggests a linear increase of the logged fuel consumption with time in the absence of maintenance actions. This amounts to an exponential growth of the added fuel consumption itself in time. However, for small values of the exponent (which is the case under usual physical conditions) the added fuel consumption can safely be linearized. After this linearization the added fuel consumption due to fouling is modeled as a sum of three contributions, $f_1 + f_2 + f_3$. The three contributions increase linearly with time with the slopes β_1, β_2 and β_3 respectively, derived from the deterioration model above. Upon performing a maintenance action the respective fouling contributions drop to zero.

There are three possible maintenance action types: 1. propeller polishing, 2. hull cleaning and 3. dry dock, corresponding to minimal, imperfect and perfect restoration degrees and having increasing maintenance costs $C_{M1} < C_{M2} < C_{M3}$ and impact on the added fuel consumption. The effects are modelled in the following “nested” manner: Action 1 reduces contribution f_1 to zero but does not affect f_2 and f_3 . Action 2 reduces only contributions f_1 and f_2 to zero. Action 3 reduces all three contributions to zero.

The optimization task is defined as follows: find the optimal scheduling of maintenance activities such that the total O&M costs (added fuel costs + maintenance costs) are minimal over a given finite planning horizon t_H , and under the following conditions:

- Any of the three actions 1,2 or 3 can be scheduled on a weekly basis.
- Some weeks $i \in T_{na}$ may be defined by the operator as unavailable for maintenance.
- Some actions $d \in A_f$ are prescheduled for specific weeks $i \in T_f$ in the future.
- The operator can set a constraint on the minimal time between actions t_{\min} .

The above formulation of the model allows us to use mixed integer linear programming in order to solve the optimization task for a given set of input parameters, $t_H, \beta_1, \beta_2, \beta_3, C_{M1}, C_{M2}, C_{M3}, C_{Fuel}$.

4.2. Optimization model and solution

We calculate the cost difference between the null policy (no maintenance during the planning horizon t_H) and a selected policy \mathbf{U} , i.e the total cost saving entailed by choosing policy \mathbf{U} . The cost saving is the objective function which we maximize. A policy is defined by a vector $\mathbf{U}=[\mathbf{U}_1 \ \mathbf{U}_2 \ \mathbf{U}_3]$ of decision variables with:

$$(15) \mathbf{U}_d = [x_{d1} \ x_{d2} \ \dots \ x_{di} \ \dots \ x_{dt_H} \ y_{d1} \ y_{d2} \ \dots \ y_{di} \ \dots \ y_{dt_H}] \quad d \in D$$

Where x_{di} are binary variables defined through

$$(15) x_{di} = \begin{cases} 1 & \text{action } d \text{ done at time } i \\ 0 & \text{otherwise} \end{cases}$$

With the constraints

$$(15) \quad \begin{aligned} x_{3i} &\leq x_{2i} \leq x_{1i} & i \in I \\ x_{1i} &= 0 & i \in T_{na} \\ x_{di} &= 1 & d \in A_f, i \in T_f \\ \sum_{j=i}^{i+t_{\min}} x_{dj} &\leq 1 & d \in D, i=1 \dots t_H - t_{\min} \\ \sum_{d=1}^3 \sum_{j=i}^{i+t_{\min}} x_{dj} &\leq 3 & i=1 \dots t_H - t_{\min} \end{aligned}$$

The decision variables y_{di} are derived from the x_{di} variables through the set of constraints:

$$(15) \quad \begin{aligned} y_{d0} + ix_{di} &\leq y_{di} \leq i & d \in D, i \in I \\ y_{di} &\leq y_{d,i+1} \leq y_{di} + (i+1)x_{d,i+1} & d \in D, i \in I \\ y_{d0} &\leq y_{d1} \leq y_{d0} + x_{d1} & d \in D \end{aligned}$$

The objective function is written in terms of the decision variables as follows:

$$(15) F = \sum_{d=1}^3 \sum_{i=1}^{t_H} [C_{Fuel} \Delta t^2 \beta_d y_{di} - C_d x_{di}] \quad \vec{C} = [C_{M1} \ C_{M2} - C_{M1} \ C_{M3} - C_{M2}]$$

The first term of the objective function amounts to the saved operational costs of fuel and the second term to the total cost of maintenance incurred by policy \mathbf{U} . Maximizing the objective function (15) with respect to the decision variables (15) under the constraints (15) and (15) yields the optimal maintenance policy \mathbf{U}^* .

Figure 3(a) shows the optimization results for $C_{Fuel} = 300\$/mT$, $C_{M1} = 10000\$$, $C_{M2} = 50000\$$, $C_{M3} = 500000\$$, $t_H = 52$ weeks and $t_{\min} = 8$ weeks. Unavailable time slots and prescheduled maintenance are marked in the plot. The resulting cost saving in this case is of 414,085\$ within the first year.

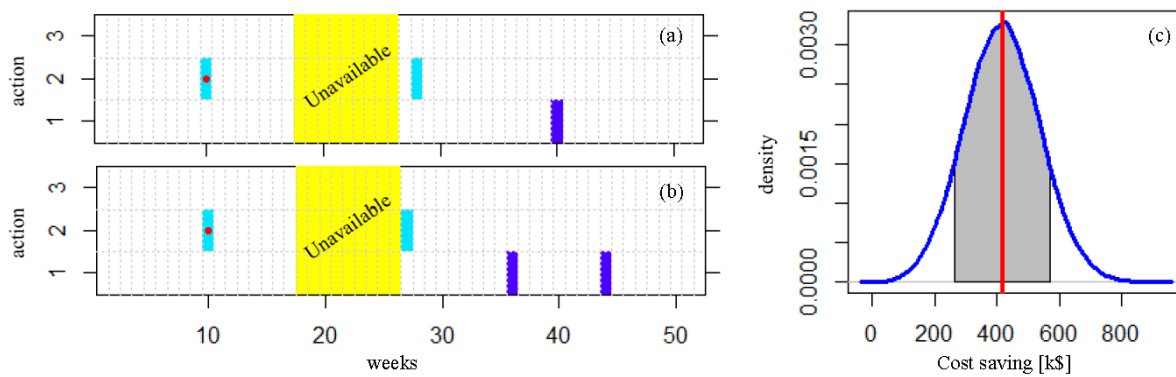


Figure 3. optimization results and uncertainty analysis. (a) The optimal policy, with 65% certainty. The red dots mark prescheduled maintenance. (b) an alternative policy, optimal with 20% likelihood. (c) probability density distribution of cost saving under the optimal maintenance schedule.

5. UNCERTAINTY ANALYSIS

In this analysis we focus on epistemic sources for uncertainty due to uncertainty of the model parameters β_1 , β_2 and β_3 (see figure 2). This uncertainty has two important implications as described below.

5.1. Uncertainty in the cost saving

In case the user implements the maintenance policy recommended by the optimization algorithm, there is an uncertainty in the cost saving due to the uncertainty in the model parameters. To estimate this uncertainty we draw sets of parameters from the posterior distribution and recalculate the cost function for each set, assuming the optimal maintenance policy \mathbf{U}^* . This yields a distribution of cost saving as depicted in figure 3(c).

5.2. Uncertainty in selecting the optimal policy

The selected optimal policy \mathbf{U}^* is affected by the model parameters. In order to assess the robustness of the recommended policy we perform a Monte-Carlo simulation of the optimization algorithm using the posterior distribution for the model parameters. This results in a set of different optimal policies which repeat with a certain likelihood. In the example above, the optimal policy \mathbf{U}^* obtained with the mean values of the posterior distribution has a certainty level of 65%. The next most probable policy is likely to occur with 20% and is depicted in figure 3(b).

6. SUMMARY

We presented a model for CBM decision making. The various ingredients of the model are all designed with a focus on practical aspects of the implementation:

- A simplified deterministic deterioration model for increased model robustness.
- Combined data-driven and physical modelling: regression model with physics-based feature selection and feature definition.
- Bayesian approach for integrating condition monitoring data with domain specific prior knowledge.
- A finite, realistic, user-defined planning horizon of the optimization framework.
- Flexible sequential optimization algorithm, including user-imposed constraints.

- Optimization considering maintenance actions with different restoration degrees: minimal, imperfect and perfect.
- Uncertainty analysis of the optimization outcomes.

The framework we present for CBM decision-making is aimed at encouraging practitioners to realize the benefits of the method and at stressing the importance of data-quality improvement. We claim that exposing new industry segments to the profits of CBM decision-making is an iterative process: Since the condition monitoring data might at first be very limited, it may often involve an initial modelling stage based largely on prior domain knowledge, possibly together with condition monitoring data which are already available at this point. The first optimization results are then expected to enhance the awareness of practitioners to the benefits of reliable condition monitoring data, leading to a higher data availability and reliability in a second step. If the model is built such that it is flexible enough to improve its accuracy upon a new data flow, the modelling and optimization results of the second step will be of a higher reliability and smaller uncertainty. This in turn, justifies the investment in condition monitoring and decision-making based on degradation modelling and optimization.

ACKNOWLEDGMENTS

The authors are grateful to Peter Fusek for very helpful discussions. This research was funded by the Swiss Commission for Technology and Innovation under Grant no. 17313.1 PFES-ES.

REFERENCES

- [1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [2] R. Ahmad and S. Kamaruddin, "A review of condition-based maintenance decision-making," *Eur. J. Ind. Eng.*, vol. 6, no. 5, pp. 519–541, 2012.
- [3] Y. G. Li and P. Nilkitsaranont, "Gas turbine performance prognostic for condition-based maintenance," *Appl. Energy*, vol. 86, no. 10, pp. 2152–2161, 2009.
- [4] R. C. M. Yam, P. W. Tse, L. Li, and P. Tu, "Intelligent predictive decision support system for condition-based maintenance," *Int. J. Adv. Manuf. Technol.*, vol. 17, no. 5, pp. 383–391, 2001.
- [5] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1161–1171, 2010.
- [6] V. T. Tran, B. S. Yang, M. S. Oh, and A. C. C. Tan, "Machine condition prognosis based on regression trees and one-step-ahead prediction," *Mech. Syst. Signal Process.*, vol. 22, no. 5, pp. 1179–1193, 2008.
- [7] M. C. Garcia, M. A. Sanz-Bobi, and J. del Pico, "SIMAP: Intelligent System for Predictive Maintenance. Application to the health condition monitoring of a windturbine gearbox," *Comput. Ind.*, vol. 57, no. 6, pp. 552–568, 2006.
- [8] A. Gelman, J. Carlin, H. Stern, D. Dunson, V. Aki, and D. Rubin, *Bayesian data analysis*. Boca Raton, FL, USA, 2014.
- [9] A. Sanchez, S. Carlos, S. Martorell, and J. F. Villanueva, "Addressing imperfect maintenance modelling uncertainty in unavailability and cost based optimization," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 1, pp. 22–32, 2009.
- [10] Z. L. Lin, Y. S. Huang, and C. C. Fang, "Non-periodic preventive maintenance with reliability thresholds for complex repairable systems," *Reliab. Eng. Syst. Saf.*, vol. 136, pp. 145–156, 2015.
- [11] F. Ren, Shuang, Gao, Yuan, Liu, "Uncertain Optimization Model for Railway Power Supply Equipments Maintenance Scheduling Problem," in *Proceedings of the Fourth International Forum on Decision Sciences, Uncertainty and Operations Research*, 2017.
- [12] R. Lu, O. Turan, and E. Boulougouris, "Voyage optimization, prediction of ship specific fuel consumption for energy efficient shipping," *Low Carbon Shipp. Conf.*, vol. 44, no. 0, pp. 1–11, 2013.
- [13] R. L. Townsin, "The Ship Hull Fouling Penalty," *Biofouling*, vol. 19, no. supplement, pp. 9–15, 2003.
- [14] Propulsion Dynamics, "A Dirty Ship is an Expensive Ship to Operate," pp. 1–11, 2016.
- [15] T. Hellström, "Optimizing Control at Sea : The Experience of the Seapacer Project," pp. 1–15, 2002.
- [16] A. F. Molland, P. S. R. Turnock, and D. A. Hudson, *Ship Resistance and Propulsion: Practical Estimation of Propulsive Power*. Cambridge university press, 2011.
- [17] L. K. Eniram, "Study of Hull Fouling on Cruise Vessels across Various Seas," 2012.
- [18] Hydrex Group, "Hull and Propeller Maintenance," *Sh. Hull Perform.*, vol. 2, no. 2, 2012.
- [19] L. B. Kulis, "Lecture 5 : Bayesian Linear Regression Bayesian Linear Regression," pp. 1–5, 2012.

Session 12

Sensor Technology and Damage Detection 3

Session Chair Dr. Erkki Jantunen

Detecting Bearing Faults Using an Ensemble Average Autocorrelation Based Stochastic Subspace Identification (*Y. Xu, P. A. Van Vuuren, X. Tang, F. Gu, A. Ball*)

Theoretical Elucidation of Pass Frequency for Multi-Flaws in a Roller Bearing and Precise Diagnosis Method Using Decision Tree and Support Vector Machine
(*Y. Mizushima, L. Song, S. Ota, P. Chen*)

Monitoring Shaft Fatigue Failures Using Online Monitoring Techniques
(*A. Abufroukh, Ahmed Onsy, Ian Sherrington*)

Detecting Bearing Faults Using an Ensemble Average Autocorrelation Based Stochastic Subspace Identification

Yuandong Xu ^a, Pieter A. van Vuuren ^b, Xiaoli Tang ^a, Fengshou Gu ^a, Andrew Ball ^a

^a Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK

^b School of Electrical, Electronic and Computer Engineering, North-West University, Potchefstroom, 2520, South Africa

ABSTRACT

Envelope analysis plays an important role in the field of bearing faults detection. Since the development of this technique, the determination of optimal bands has been a prior challenge. Fast Kurtogram (FK) is an outstanding approach to select an optimal band for further analysis; however, fast Kurtogram is not robust enough to withstand the influence of white noise and large aperiodic impulses. Hence, a more robust method is introduced to extract the narrow bands for envelope analysis, which is ensemble average autocorrelation based stochastic subspace identification (SSI). The detector performs well in denoising and highlighting the periodic impulses owing to the outstanding characteristics of autocorrelation function and stochastic subspace identification. Considering the results of simulation study and experimental evaluation, it can be concluded that the proposed method is more effective and robust to detect bearing faults than fast Kurtogram.

Keywords: fault detection, ensemble average autocorrelation, SSI.

Corresponding author: Yuandong Xu (Yuandong.xu@hud.ac.uk)

1. INTRODUCTION

Bearings have been widely used since the invention of the wheel and they play an important role in the field of machinery. The failure of bearings may result in the breakdown of machines or even catastrophic accidents. Considering the safety and efficiency of machines, condition monitoring of bearings is significant. Vibration signals induced by the faults have been investigated for decades, which is a powerful technique to monitor the working conditions of bearings. The high frequency resonance technique (HFRT), later called “envelope analysis”, was developed [1] owing to the outstanding ability of the good resolution after the frequency shift from high carrier frequency bands to low fault frequency bands. Since the technique of high frequency demodulation was introduced by Darlow, there has long been a discussion on how to choose the optimum bandwidth for the demodulation associated with envelope analysis[2]. As Antoni [3] studied spectral kurtosis (SK) thoroughly, Fast Kurtogram [4] based on short-time fast Fourier transform (STFT) and wavelet transform (WT) has been developed and explored by many researchers[5], [6]. Gu [7] introduced modulation signal bispectrum (MSB) to identify and quantify the common faults of a compressor. Tian [8] as well as Rehab [9] et al verified modulation signal bispectrum (MSB) with high performance of robustness to detect the optimal bands and bearing faults even though the signal-to-noise ratio (SNR) is very low. In this paper, a novel method, ensemble average autocorrelation based stochastic subspace identification (SSI) is developed to automatically select the optimal bands according to the characteristics of modulation signals and detect the bearing faults based on the outstanding features of autocorrelation functions.

Fast Kurtogram has been widely used owing to the effectiveness of optimal narrowband selection and it is generally accepted as a powerful method to select the optimum frequency band for envelope analysis. Hence, Fast Kurtogram is used as a benchmark to examine the performance of the novel method. This paper is structured as follows. Firstly, section 2 summarizes the main steps of the fault detection algorithm used in this paper. The robustness of the detection algorithm for disturbance signals is investigated by means of

simulation in section 3. Section 4 briefly compares the performance of the new fault detection algorithm with the fast Kurtogram on a laboratory test-bench, while concluding remarks are made in section 5.

2. ENSEMBLE AVERAGE AUTOCORRELATION BASED SSI

A defective bearing signal is a typically amplitude modulation signal [10]. The vibration signal of a bearing system with a local fault consists of periodical impulses, system resonance and noise. Hence, it can be expressed as [10]

$$x(t) = x_f(t) \cdot x_q(t) \cdot x_{bs}(t) + x_s(t) + n(t) \quad (16)$$

Where, $x_f(t)$ is the periodical impulses induced by the local defect, which occurs periodically corresponding to the shaft rotation; $x_q(t)$ is the modulation effect produced by the non-uniform load distribution; $x_{bs}(t)$ is the bearing resonant behaviour; $x_s(t)$ is the machine vibration induced by the machinery dynamics; $n(t)$ is the inevitable noise which results from the working environments and the data acquisition system. It is obvious that the vibration signal of faulty bearings is modulated and rich in noise, therefore demodulation and denoising are essential to detect bearing faults effectively. Ensemble average autocorrelation [11]–[13] based stochastic subspace identification [14]–[17] (EAAC-SSI) is insensitive to noise effect and able to determine an optimal central frequency for demodulation analysis. The main steps of EAAC-SSI in finding the frequency are shown in figure 1.

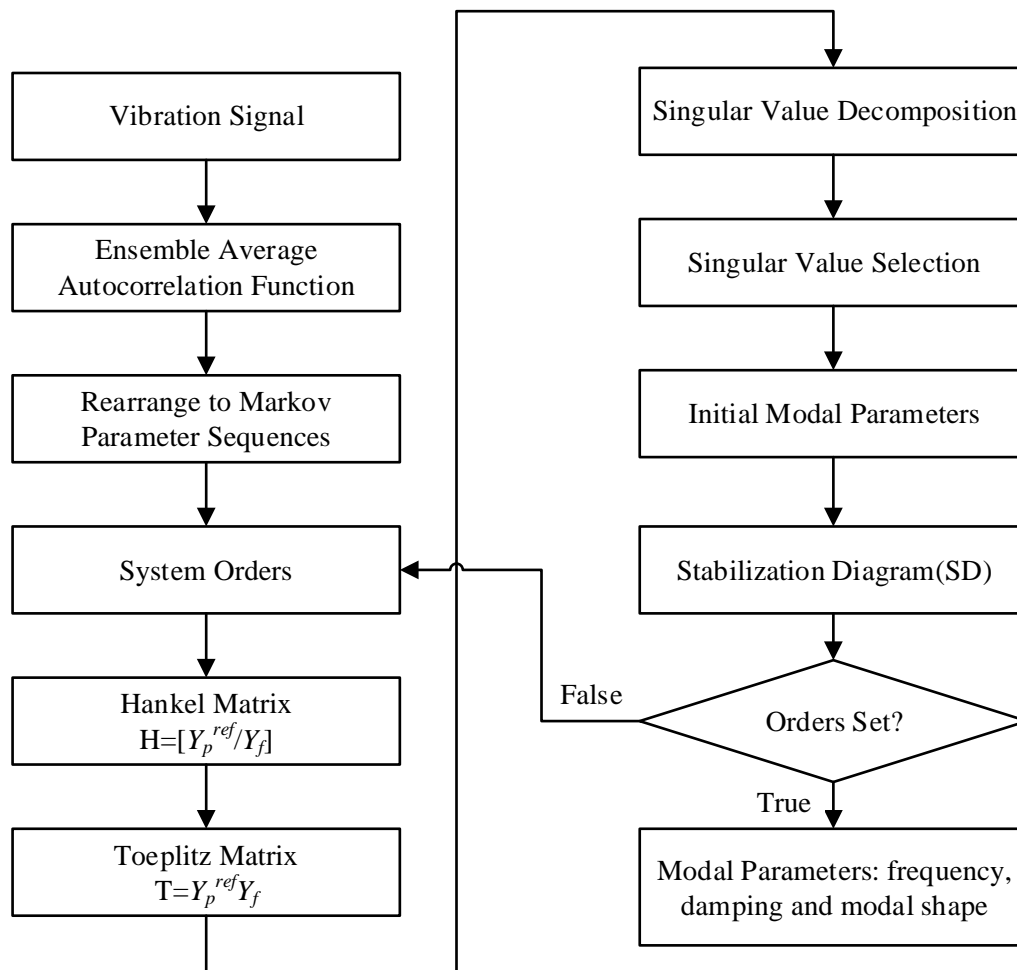


Figure 1. Flow chart of EAAC-SSI

3. SIMULATION STUDY

In order to examine the robustness of two methods, two cases based on white noise and aperiodic impulses are investigated. In this section an artificially generated bearing vibration signal is used. The signal represents an idealised acceleration vibration signal consisting of rapidly decaying exponential pulses multiplied with a high frequency sinusoid. An example of the simulated signal is shown in the top graph in figure 2. The key parameters of the simulated signal are given in table 1.

Table 1: Key parameters of simulated signals

| Parameters | Value |
|------------------------------|----------|
| Sampling Frequency F_s | 25600 Hz |
| Fault Frequency f_f | 89.2 Hz |
| Resonance Frequency f_{bs} | 5000 Hz |
| Rotating Frequency f_r | 25 Hz |

3.1. Case of white noise

In this case, white noise was added to the simulated signal and figure 2 depicts the simulated time waveform of the fault vibration signal and the added high level noise.

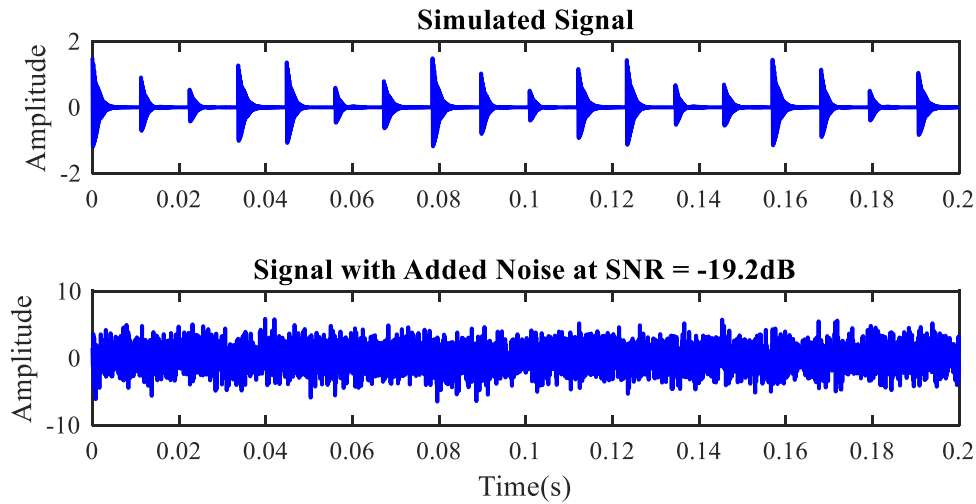


Figure 2. Simulated signal with added white noise

Both fault detection methods, namely EAAC-SSI and fast Kurtogram were applied to the simulated signal. The results of these two methods are shown in figure 3 and it is clear that EAAC-SSI is able to detect the fault signature while fast Kurtogram fails to highlight the fault frequency due to the influence of white noise.

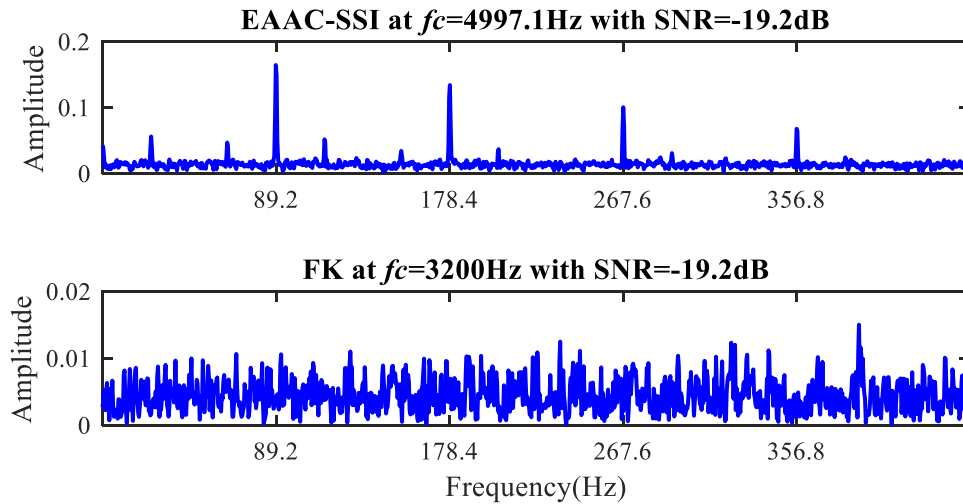


Figure 3. Results of two methods processing white noise signal

As the increase of SNR value, the signal was submerged in the white noise, which makes fast Kurtogram is invalid to determine the proper frequency bands. Based on the central frequencies obtained by two approaches, the bandwidth for demodulation is 900 Hz. When the signal to noise ratio is equal to -19.2 dB, the high level noise makes the kurtosis spectra unable to indicate the non-Gaussian components and locate the frequency bands. While ensemble average autocorrelation based SSI highlights the fault frequency and corresponding sidebands due to its outstanding noise suppression abilities.

3.2. Case of aperiodic impulses

As shown in the previous section, the two methods (EAAC-SSI and FK) differ in their ability to perform fault detection in the presence of white noise. In practice, bearing vibration signals are also contaminated with impulse noise. In this section, the bearing signal with several impulses is generated to evaluate the novel method and the benchmark one.

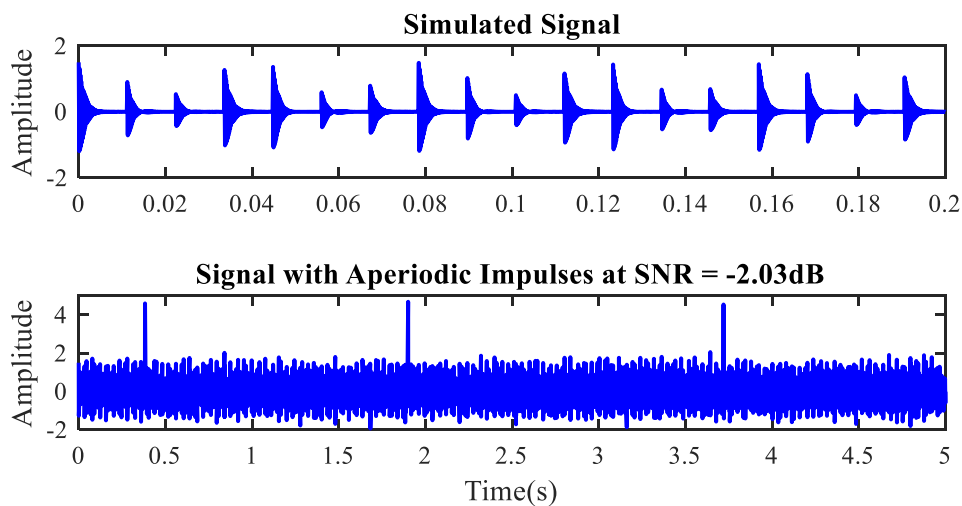


Figure 4. Simulated signal with aperiodic impulses

The noise-free signal is same with the previous one in the first case, the details of which are listed in table 1. The lower subgraph in figure 4 partially shows that the modulation signal is interfered by 20 random impulses and low level Gaussian noise in the time domain. The amplitude of the aperiodic impulses are the maximum value of the white noise and the positions of the impulses are determined by the normally distributed random function in Matlab.

As bearing fault signal with high signal to noise ratio is complete, two methods are applied to fault detection and the results are demonstrated in figure 5.

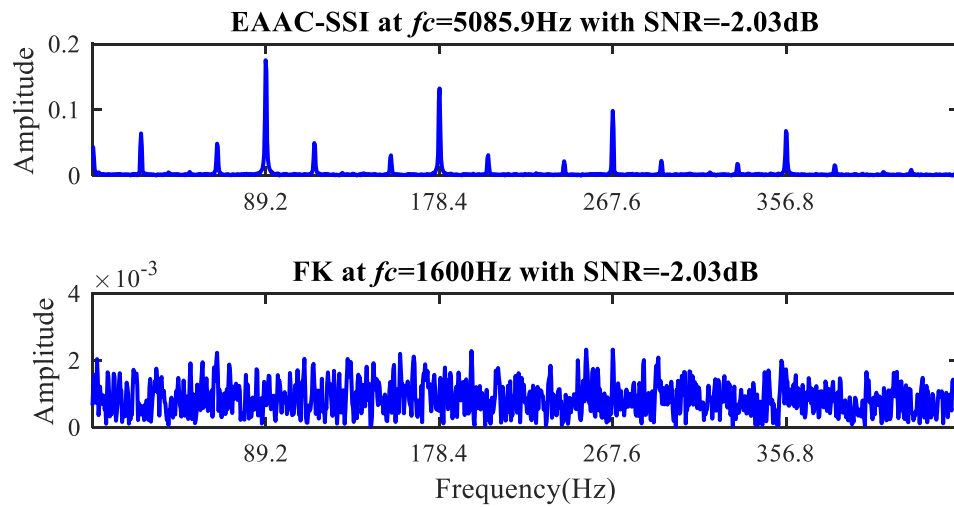


Figure 5. Results of two methods processing aperiodic impulses signal

The central frequencies of EAAC-SSI and FK are 5085.9 Hz and 1600 Hz respectively and further envelope analysis with bandwidth 900 Hz referred to the determined frequencies was carried out. The fault frequency 89.2 Hz and its harmonics are obvious in the first subgraph and additionally, the rotating frequency and the sidebands are apparent. However, FK could not locate the optimal band due to the influence of impulses which leads to the high levels of interference. Hence, FK fails to detect the bearing faults under these circumstances

4. EXPERIMENTAL STUDY

In order to evaluate the performance of the two detector algorithms, the motor bearing with a local defect on the inner race is installed to collect the vibration signal. The layout of the test rig is described in figure 6(a) and the fault on the inner race is illustrated in figure 6(b).

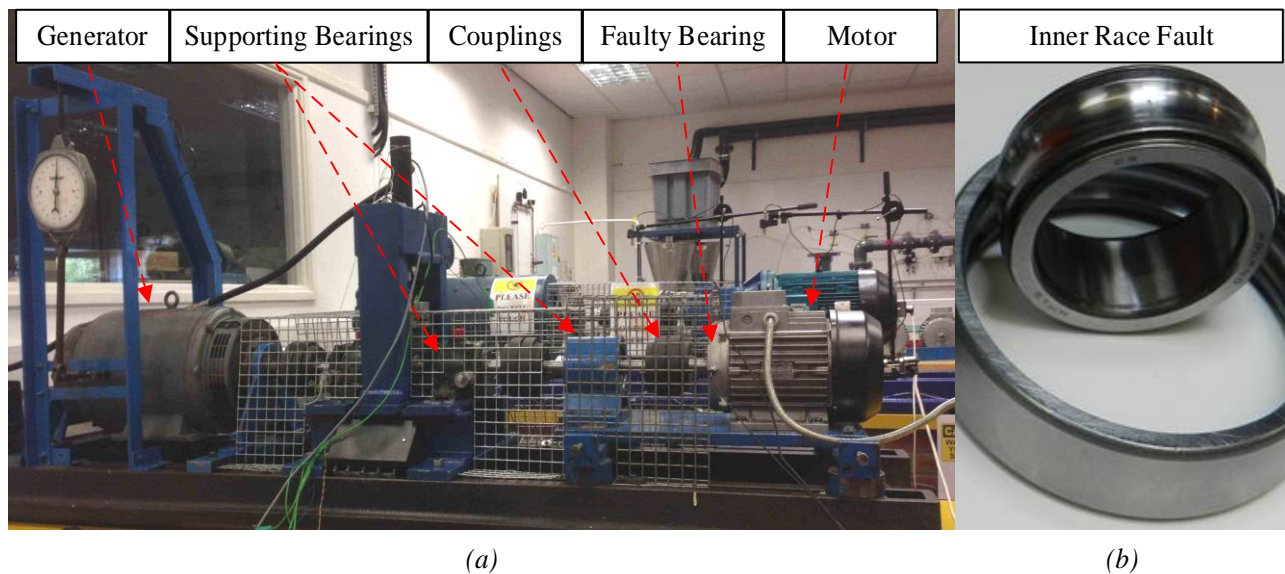


Figure 6. Layout of the motor bearing test rig

The specification of the faulty bearing is shown in table 2.

Table 2: Specification of the ball bearing

| Rolling Ball Bearing | Value |
|-----------------------|---------|
| Number of Balls | 9 |
| Ball Diameter | 9.53 mm |
| Pitch Circle Diameter | 46.4 mm |
| Contact Angle | 0 |

When the test rig was operated at 100% speed (1450rpm) and 100% torque load, the vibration signal was acquired once steady state operation was achieved. In total 30 seconds of vibration data was logged at a sampling frequency of 96 kHz.

Referred to the specifications and the practical rotating speed, the theoretical fault frequency (f_i) is 128.2 Hz and the analysis results of two methods are illustrated in figure 7.

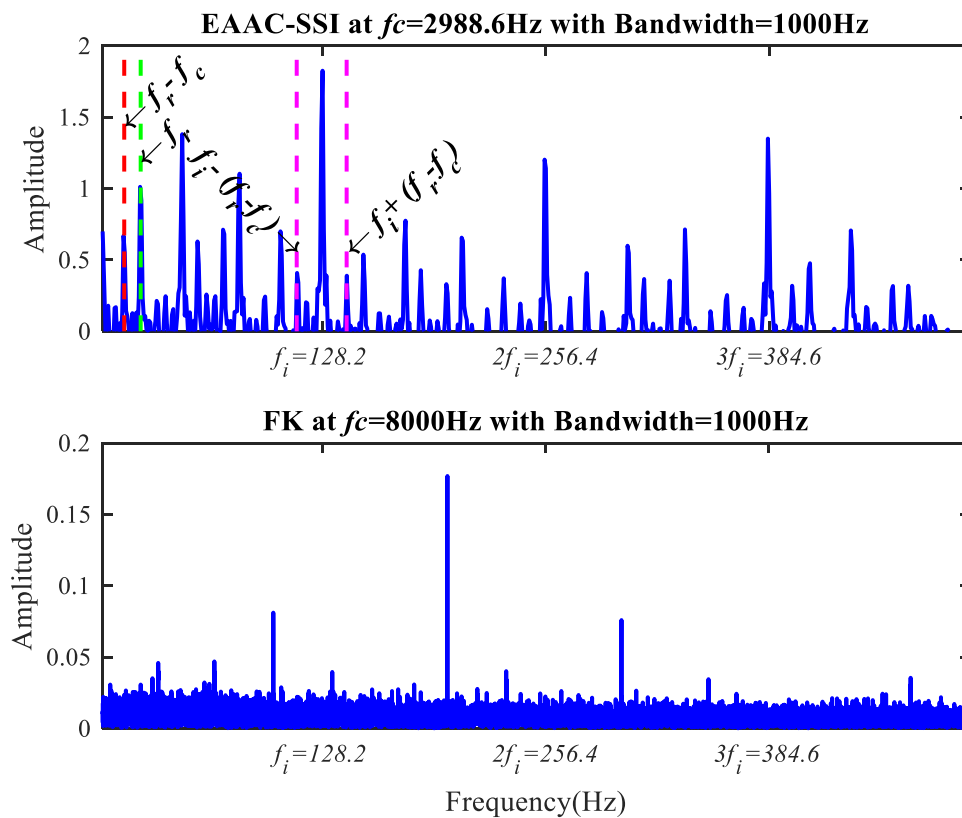


Figure 7. Results of two method processing ball bearing vibration signal

The demodulation of EAAC-SSI demonstrates the inner race fault and the corresponding harmonics. In addition, the cage frequency (f_c) is also apparent in the spectrum. Practically, the fault on the inner race was made manually and during the disassembly and reassembly, the cage of the bearing is not as perfect as the initial condition. Hence, the cage fault frequency is detected by the novel method. Nevertheless, the central frequency 8000 Hz was selected by FK and it cannot extract the fault signature due to the improper frequency band.

5. CONCLUSIONS

In order to examine the robustness of the ensemble average autocorrelation based stochastic subspace identification, the simulation study shows the proposed method performs robustly under the influence of white noise and aperiodic impulses. In the experimental evaluation, the benchmark method, Fast Kurtogram, failed to determine the optimal band while the novel method determined the optimal band and highlighted the inner race fault frequency and the cage fault frequency. Both simulation and experimental results demonstrate that compared with the performance of Fast Kurtogram, the proposed method is more effective and more robust.

REFERENCES

- [1] M. S. Darlow, R. H. Badgley, and G. W. Hogg, 'Application of High-Frequency Resonance Techniques for Bearing Diagnostics in Helicopter Gearboxes.', No. MTI-74TR25. MECHANICAL TECHNOLOGY INC LATHAM NY, Oct. 1974.
- [2] R. B. Randall and J. Antoni, 'Rolling element bearing diagnostics—A tutorial', *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.
- [3] J. Antoni, 'The Spectral Kurtosis: A Useful Tool for Characterising Non-Stationary Signals', *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 282–307, Feb. 2006.
- [4] J. Antoni and R. B. Randall, 'The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines', *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, Feb. 2006.
- [5] F. Gu, X. Tian, Z. Chen, T. Wang, I. Rehab, and A. Ball, 'Fault severity diagnosis of rolling element bearings based on kurtogram and envelope analysis', 2014.
- [6] Y. Lei, J. Lin, Z. He, and Y. Zi, 'Application of an improved kurtogram method for fault diagnosis of rolling element bearings', *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1738–1749, Jul. 2011.
- [7] F. Gu, Y. Shao, N. Hu, B. Fazenda, and A. Ball, 'Motor current signal analysis using a modified bispectrum for machine fault diagnosis', in *2009 ICCAS-SICE*, 2009, pp. 4890–4895.
- [8] X. Tian, F. Gu, I. Rehab, G. Abdalla, and A. Ball, 'A robust fault detection method of rolling bearings using modulation signal bispectrum analysis', presented at the 28th International Congress of Condition Monitoring and Diagnostic Engineering Management (COMADEM 2015), Buenos Aires, Argentina, 2015.
- [9] I. Rehab, X. Tian, F. Gu, and A. Ball, 'The fault detection and severity diagnosis of rolling element bearings using modulation signal bispectrum', presented at the Eleventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 2014.
- [10] Y.-F. Wang and P. J. Kootsookos, 'Modeling of Low Shaft Speed Bearing Faults for Condition Monitoring', *Mech. Syst. Signal Process.*, vol. 12, no. 3, pp. 415–426, May 1998.
- [11] Z. D. Bi, Y. Chen, Z. Z. Peng, and Y. Zhang, 'The Demodulating Principle of Autocorrelation-Envelope and its Application to Gearbox Diagnosis', *Appl. Mech. Mater.*, vol. 380–384, pp. 1029–1034, 2013.
- [12] W.-Y. Wang and M. J. Harrap, 'Condition monitoring of ball bearings using an envelope autocorrelation technique', *Mach. Vib.*, vol. 5, no. 1, pp. 34–44, 1996.
- [13] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. Berlin ; London: Springer, 2006.
- [14] A. Dahlén, 'Identification of Stochastic Systems: Subspace Methods and Covariance Extension', 2001.
- [15] P. G. Bakir, 'Automation of the stabilization diagrams for subspace based system identification', *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14390–14397, Nov. 2011.
- [16] G. Zhang, B. Tang, and G. Tang, 'An improved stochastic subspace identification for operational modal analysis', *Measurement*, vol. 45, no. 5, pp. 1246–1256, Jun. 2012.
- [17] E. Mrabet, M. Abdelghani, and N. Ben Kahla, 'A New Criterion for the Stabilization Diagram Used with Stochastic Subspace Identification Methods: An Application to an Aircraft Skeleton', *Shock Vib.*, vol. 2014, 2014.

Theoretical Elucidation of Pass Frequency for Multi-flaws in a Roller Bearing and Precise Diagnosis Method Using Sequential diagnosis by Support Vector Machine

Yuto Mizushima, Liuyang Song, Shyohei Ota and Peng Chen¹

¹ Graduate School of Bioresources, Mie University
1577 Kurimamachiya-cho, Tsu-shi, Mie, 514-8507, Japan

ABSTRACT

Roller bearing is an important part in rotating machinery, which supports rotating shaft in the machine. Accordingly, it is important to monitor the condition of roller bearing in various operations for preventing the being damages. The fault types that often occur in the roller bearing are out race flaw, inner race flaw and roller element flaw. Although the theory and method for diagnosing one flaw in the roller bearing have been established, the precise diagnosis method for identifying the locations of multi-flaws in a roller bearing has not been clarified theoretically yet. Therefore, in this study, the method for precisely diagnosing multi-flaws in a roller bearing is proposed as follows:

Firstly, it is proved theoretically that the pass frequency used for diagnosing single flaw can be used for diagnosing multi-flaws. Moreover, in order to confirm the accuracy of theory, the spectrum of the artificial envelope waveform for the bearing multi-flaws by simulation is compared with that of the data obtained from experiments. It was found that the pass frequencies caused by the bearing multi-flaws are same in both the cases of theory and the experiments. Secondly, intelligent diagnosis method based on support vector machine (SVM) is proposed for automatic diagnosis. The dedicated symptom parameter are not required when using SVM. This is an advantage when the proposed method is applied for on-line diagnosis. Finally, the sequential diagnosis method is proposed for discriminating the conditions of bearing, such as normal or abnormal, outer race flaw or others, inner race flaw or roller element flaw.

The efficiency of the method is verified by practical fault diagnosis of bearings with multi-flaws. The bearings used in the experiments have scratches on race or roller element. The number and positions of the scratches are different.

Key words: Roller Bearing, Multi-flaws, Symptom parameter, Sequential diagnosis, Support vector machine

Corresponding author: P. Chen (chen@bio.mie-u.ac.jp)

1. INTRODUCTION

Although, there are many methods for fault diagnosis of roller bearing at early stage (single-flaw) have been proposed[1], the fault diagnosis method for middle stage or final stage, such as multi-flaws, have not been established. It is considered that fault detection and bearing replacing at early stage is best. However, when fault detection is missed at the early stage, the fault level will worsen into the middle stage (multiple scratches and flaking). At the middle stage the fault of bearing have an exponential progress. If fault diagnosis cannot be carried out as soon as possible, facility accidents due to the destruction of bearings may occur. Bearing fault diagnosis includes fault detection and precise diagnosis. The purpose of fault detection is to distinguish normal and abnormal. The objective of precise diagnosis is to judge the fault types. When precise diagnosis is conducted, pass frequency is also used for indicating the location of the fault. There is a theoretical description about pass frequency applied for bearing fault diagnosis of single-flaw. However, there is no description stating that it can be applied for the fault diagnosis of bearing with multiple scratches [3] - [5].

In this paper, it is proved in theory that the pass frequency can be used for bearing fault diagnosis with multi-flaws by Fourier transform. Then, an intelligent bearing fault diagnosis method based on support vector machine (SVM) and sequential diagnosis method is proposed. The effectiveness of the proposed method is confirmed by a rotating machine. In addition, when the classification method such as decision tree and neural network is used for precise diagnosis, the parameters of the bearing such as outer diameter, inner diameter, roller element diameter, number of rolling elements etc., are necessary for calculating the symptom parameters. However, when on line diagnosis is conducted in an actual factory, the specification of the bearing is unknown because the machine has not been disassembled. In this case, it is impossible to calculate the pass frequency. Moreover, it is not easy to find symptom parameter because there are so many different parameters for discriminating different fault types. When the proposed method is used for precise diagnosis, the specification of bearing is not required. It is an intelligent and automated fault diagnosis method that can be performed by computer. Finally, the effectiveness of the method is verified by the results of the experiments on a rotating machine.

2. THEORETICAL CONSIDERATION

As mentioned, it is unknown whether pass frequency calculated for condition diagnosis of single-flaw can be used for multi-flaws. Therefore, this chapter proved that condition diagnosis of multi-flaws could be carried out by using the pass frequency for single-flaw condition diagnosis through the theoretical demonstration of Fourier Transform and the analysis of artificial envelope waveform.

2.1. Leading in Fourier transform

In order to theoretically verify the contention of this paper, the signal for bearing fault diagnosis is regarded as a periodic waveform. The periodical artificial envelope waveform are processed by Fourier transform according to Eq. (1) ~ (4) [6]. Firstly, the pass frequency of the bearing with single-flaw is calculated. Then, the pass frequency of the bearing with multi-flaws is calculated. It is proved that the pass frequencies of the two cases are the same.

$$G(f) = \int_{-\infty}^{\infty} g_f(t) e^{-2\pi i f t} dt \quad (1)$$

$$g_f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi n t}{T}\right) + b_n \sin\left(\frac{2\pi n t}{T}\right) \right] \quad (2)$$

$$a_n = \frac{2}{T} \int_0^T g(t) \cos\left(\frac{2\pi n t}{T}\right) dt \quad (3)$$

$$b_n = \frac{2}{T} \int_0^T g(t) \sin\left(\frac{2\pi n t}{T}\right) dt \quad (4)$$

$g(t)$ is the periodic waveform (envelope waveform), $G(f)$ is the spectrum of $g(t)$ after the Fourier transform.

2.1.1 About single-flaw pass frequency

Firstly, the discussion of the case with single-flaw is commenced in this paragraph. The artificial envelope waveform will be simulated and the periodic function will be defined. Then, the envelope spectrum can be obtained after the Fourier transform.

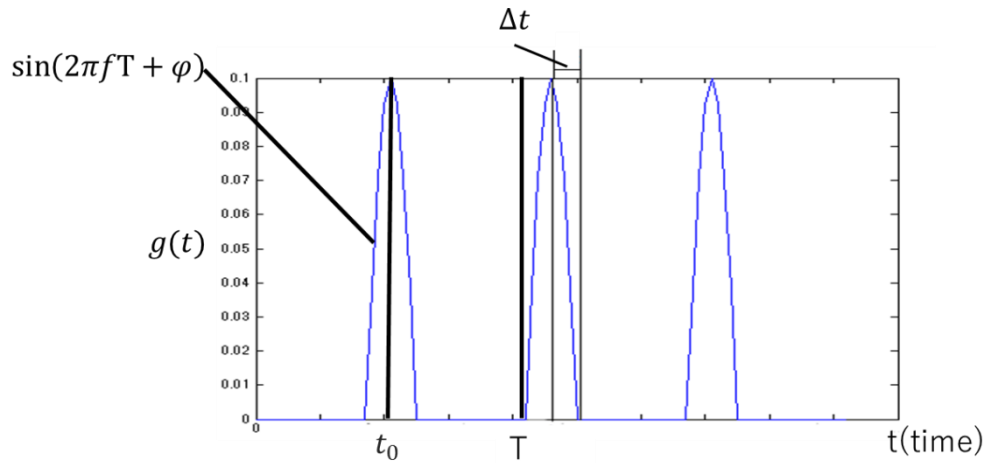


Figure 1. Artificial waveform

As shown in figure 1, the artificial periodical waveform of bearing fault diagnosis is defined by Eq. (5). The waveform with only one impact in one period is used to simulate the signal of the bearing with single-flaw.

$$g(t) = \begin{cases} 0 & (0 \leq t \leq t_0 - \Delta t) \\ \sin(2\pi f t + \varphi) & (t_0 - \Delta t \leq t \leq t_0 + \Delta t) \\ 0 & (t_0 + \Delta t \leq t \leq T) \end{cases} \quad (5)$$

The Fourier transform is taken according to Eq. (3)-(4). The result is shown as follows:

$$Gs(f) = \frac{1}{2} \{ p(\delta(f - nf_0) + \delta(f + nf_0)) - qi(\delta(f - nf_0) + \delta(f + nf_0)) \} \quad (6)$$

Where, $f_0 = 1/t_0$ is the pass frequency, and $\delta(0) = \infty$ in theory. The result shows there are peaks in the spectrum at the frequencies of $f = nf_0$.

2.1.2 About multi-flaws pass frequency

The same processing is carried out for the case of multi-flaws. The simulation waveform is constructed by Eq. (7). The result after the Fourier transform is shown in Eq. (8).

$$g(t) = \begin{cases} 0 & (0 \leq t \leq t_1 - \Delta t_1) \\ A1 \sin(\omega_1 t + \varphi_1) & (t_1 - \Delta t_1 \leq t \leq t_1 + \Delta t_1) \\ 0 & (t_1 + \Delta t_1 \leq t \leq t_2 - \Delta t_2) \\ A2 \sin(\omega_2 t + \varphi_2) & (t_2 - \Delta t_2 \leq t \leq t_2 + \Delta t_2) \\ 0 & (t_2 + \Delta t_2 \leq t \leq T) \end{cases} \quad (7)$$

$$G(f) = \frac{1}{2} \{ X(\delta(f - nf_0) + \delta(f + nf_0)) - Yi(\delta(f - nf_0) + \delta(f + nf_0)) \} \quad (8)$$

The result shows that peaks appear in the spectrum at the frequencies of $f = nf_0$. It is proved that the positions of the peaks in the spectrum of the two cases (single-flaw and multi-flaws) are the same. Addition, even if the number of scratches increased, just the coefficient X and Y in Eq. (8) will change, the peaks will still at the same positions.

3. INTELLIGENT DIAGNOSIS

3.1. Support vector machine (SVM)

The basic idea of Support Vector Machine (SVM) [7] is that when two classes exist in one area, SVM can automatically and optimally decide the border and divide the area into two classes, as shown in figure 2. The data in one class which is closest to the other class is used as a support vector. In figure 3, the circled data are regarded as support vectors, and used to draw the identification border with the maximum Euclid distance. The goal of SVM is to make the margin between the support vectors and the identification border as much as possible. This is also the most significant advantage of SVM.

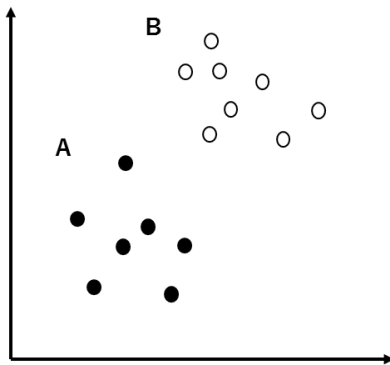


Figure 2. Data belonging to 2-classes

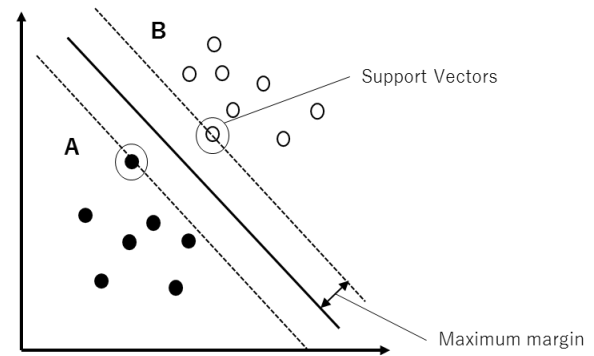


Figure 3. Max Euclid distance border

3.2. Intelligent diagnosis based on SVM

In this paper, the intelligent fault diagnosis method based on SVM is proposed and combined with the optimal symptom parameters. The training data which state is known is prepared. The linear division is then carried out by SVM to get the maximum margin. Finally, the test data can be classified and the condition of the bearing can be recognized by the trained SVM. Besides, the optimal condition monitoring and fault diagnosis of the machine are also possible.

4. EXPERIMENT

Firstly, the experimental data was measured in each state. The machine used in the experiment is shown as figure 4. The bearings faults were artificially introduced as shown in figure 4. The vibration data was obtained by two acceleration sensors set in the vertical direction and horizontal direction. Taking the machinery used in actual plant into consideration, static load (400 kg) was acted on the shaft in the experiment. The measuring time was 10 seconds and sampling frequency was 100 kHz. The shaft was rotated at 600 rpm, 900 rpm and 1200 rpm. The states of the bearings were normal, outer race flaw, inner race flaw and roller element flaw, and there were 6 types of single-flaw and multi-flaws in each fault state.

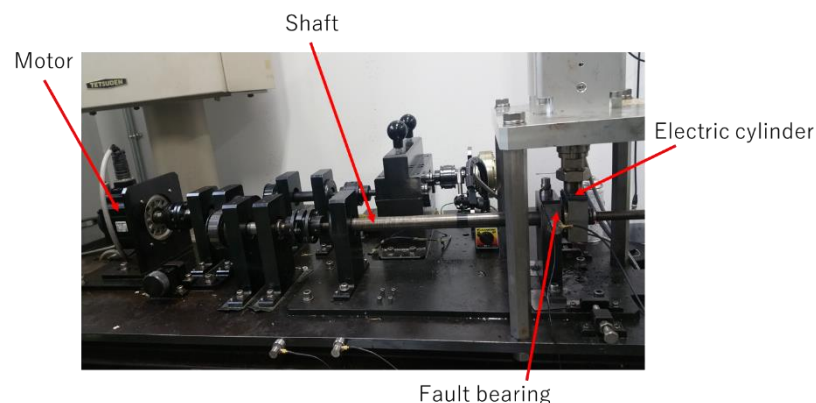


Figure 4. The rotating machinery used in the experiment



Figure 5. The bearings with artificial damage. Left to right outer race fault, inner race fault, roller element fault

4.1. Election of optimal parameter by DI

6 dimensionless symptom parameters (SPs) in the frequency domain were calculated. The DI value of each SPs were also obtained according to Eq. (9). Then, two sensitive SPs with larger DI values were selected. The data were plotted with the two selected SPs as the vertical axis and the horizontal axis. Finally, classification analysis was performed in two-dimensions.

$$DI = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad (9)$$

4.2. The sequential diagnosis by SVM

Figure 6 shows an example of classification analysis by SVM. The support vector is surrounded by small circles. In addition, the test data was also prepared to confirm whether classification after training is correct. The test samples at each state that enclosed by large circles were respectively input into the trained SVM. In this research, in order to improve the accuracy, 4 states were not identified by single matching. Normal, outer race flaw, inner race flaw, and roller elements flaw were identified by sequential diagnosis method as shown in figure 7. In the training step, the symptom parameters (SPs) are calculated and the sensitive SPs are selected by DI. Then, SVM will be trained by the data which states are known. The identification borders are also drew sequentially for fault diagnosis. In the classifying step (diagnosis step), the condition of the bearing will be discriminated by using the identification borders. Firstly, whether the bearing is in normal or abnormal state will be judged. Secondly, it is judged whether in the outer race flaw state or not. Finally, it is judged whether in the inner race flaw state or roller element flaw.

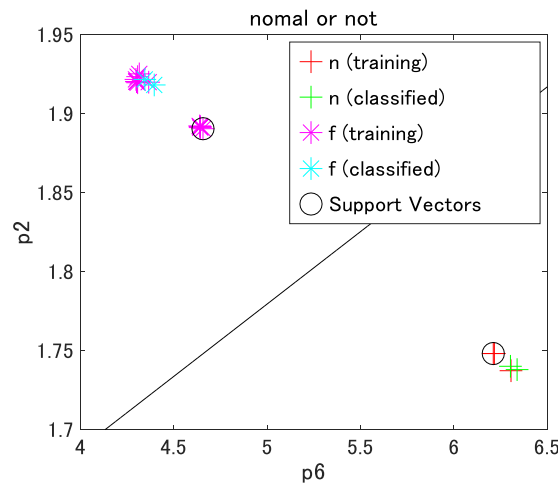


Figure 6. The classification by SVM

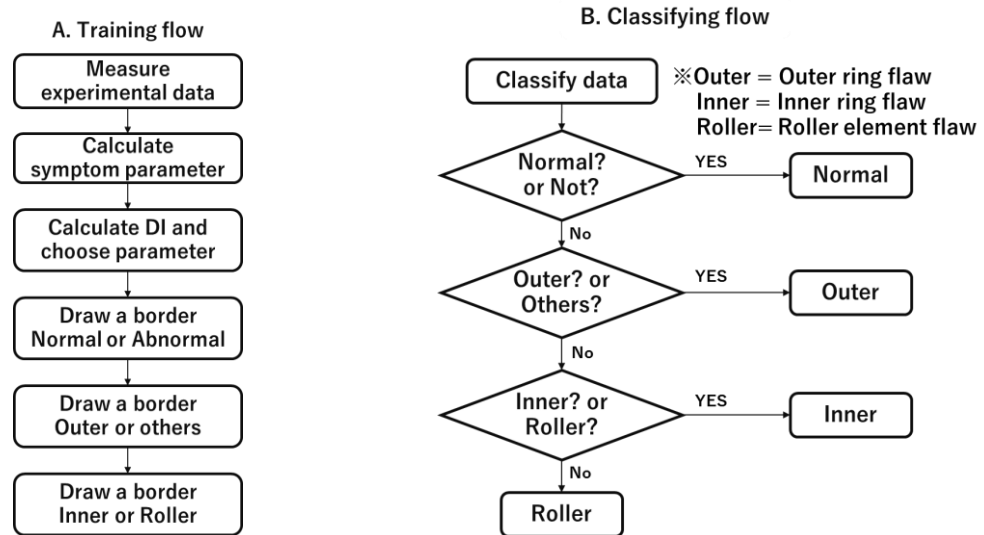


Figure 7. Flow chart of Sequential diagnosis

4.3. Result of classification

The classification results of the experiment are shown in figure 8. Fault detection, diagnosis of outer race flaw, diagnosis of inner race flaw, and diagnosis of roller element flaw are sequentially performed from left to right. The states are correctly discriminated. The effectiveness of the proposed method are verify by the results. Also, the results of other data are summarized in table 1 and the accuracies of the classification are shown.

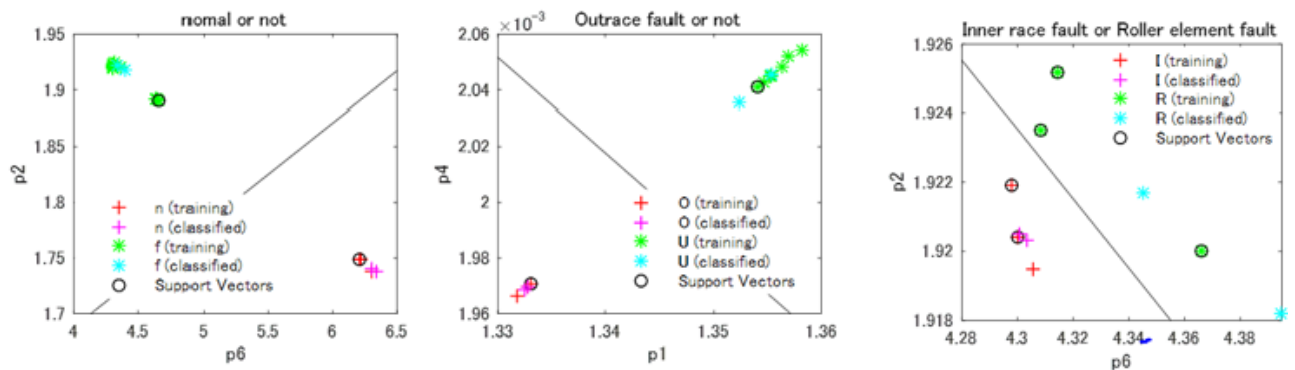


Figure 8. The result of sequential diagnosis by SVM

| Table 1 The result of classification | | | | | | |
|--------------------------------------|------|------|------|------|------|------|
| | NO.1 | NO.2 | NO.3 | NO.4 | NO.5 | NO.6 |
| Simple diagnosis | 100% | 100% | 100% | 100% | 100% | 100% |
| Outer ring | 100% | 100% | 100% | 100% | 100% | 100% |
| Inner ring | 100% | 100% | 100% | 100% | 100% | 100% |
| Roller element | 100% | 100% | 100% | 100% | 100% | 100% |

5. CONCLUSION

In this paper, whether the pass frequency of bearing with single-flaw can be used for bearing fault diagnosis with multi-flaws was theoretically clarified by using Fourier transform. Then, an intelligent bearing fault diagnosis method was proposed which is based on support vector machine (SVM) and sequential diagnosis method. The performance of method is verify by fault diagnosis using the bearings with plural scratches on outer race, inner race and roller elements. The classification accuracies of the data from experiments are 100%. In the future, in order to clarify the universality of the intelligent fault diagnosis method based on SVM, data under various conditions will be acquired for precision diagnosis. In addition, nonlinear classification will be realized so that classification analysis can be performed even the distribution of the data is complicated.

REFERENCES

- [1] Y. Yang, D. Yu And J.Cheng: A Fault Diagnosis Approach For Roller Bearing Based On Imf Envelope Spectrum And Svm
- [2] Burgess, C. J. C. (1998). A Tutorial On Support Vector Machines For Pattern Recognition. *Data Mining And Knowledge Discovery*, 2, 955–974.
- [3] Muhammet Unal, Mustafa Demetgul, Mustafa Onat And Haluk Kucuk, Fault Diagnosis Of Rolling Bearing Based On Feature Extraction And Neural Network Algorithm; *Recent Advances In Telecommunications, Signals And Systems*.
- [4] I.M.Howard, A Review Of Rolling Element Bearing Vibration “Detection, Diagnosis And Prognosis”; *Defence Science And Technology Organisation*, October 1994
- [5] Shahab Hasanzadeh Ghafari, A Fault Diagnosis System For Rotary Machinery Supported By Rolling Element Bearings; 2007
- [6] Fredric J. Harris: *Windows, Harmonic Analysys, And The Discrete Fourier Transform*, 1976
- [7] J.A.K.Suykens And J.Vandewalle, Least Squares Support Vector Machine Classifiers; *Neural Processing Letters* 9:293-300, 1999

Monitoring Shaft Fatigue Failure Using an Online System

A. Abufroukh, A. Onsy, I. Sherrington

Jost Institute for Tribotechnology, University of Central Lancashire, Preston PR1 2HE, UK

ABSTRACT

Shafts are used in almost all mechanical applications where rotational movement is transferred between machine elements. However, shafts are subjected to material defects, physical damage and fatigue which may cause catastrophic failure. The four main causes of shaft failure are: corrosion, wear, fracture due to overload and fatigue. Fatigue is common and often occurs when a varying load is applied over many cycles, with the subsequent changes in stress leading to cracks or fractures. This paper will outline some of the development of a condition monitoring system for identifying developing fatigue failure in shafts. The aim is to develop a system using sensors and online monitoring approaches with predictive or proactive maintenance strategies to facilitate the development of a failure detection system. The development of a test apparatus to detect developing fatigue failure in shafts is described in the paper.

Keywords: Fatigue failure, condition monitoring, fracture, catastrophic failure, online active systems

Corresponding authors: Ahmed N-s Abufroukh (anabufroukh@uclan.ac.uk)

1. INTRODUCTION

Shafts are used in many mechanical applications where rotational movements are transmitted between elements. Shafts experience many types of failure, which may be caused by any one of a number of different factors. It is, therefore, often difficult to determine the precise cause of failure. However, depending on the rate of failure, shaft deterioration may lead to a catastrophic failure resulting in personal injury, machine damage or environmental contamination. The typology of failures that commonly occur in shafts can be classified as: corrosion, wear, overload and fatigue [1, 2, 3]. Corrosion and wear tend to cause catastrophic failure less frequently than overload or fatigue as they develop slowly and leave more obvious evidence of their development. Overload failures occur when applied forces exceed the yield strength or the tensile strength of the component [1, 4]. The appearance of the developing damage in this type of failure depends on whether the material is brittle or ductile – as depicted in figure 1 below.

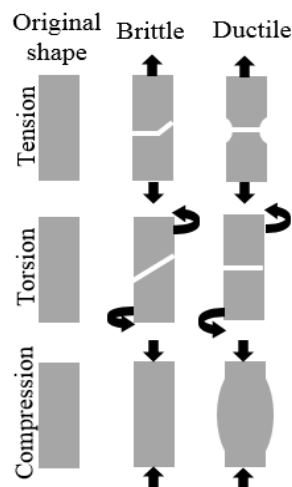


Figure 1. The appearance of tensile, torsional and compressive shaft overload failure (After Sachs [1])

If a brittle shaft undergoes excessive tension, cracks are initiated and spread very quickly with little deformation evident at the surface and the shaft will separate. In a ductile material, failure occurs after

plastic deformation develops and the load continues to increase such that the component can no longer support the applied force. The component will resist any additional extension unless there is an increase in the applied stress. Figure 2 illustrates plastic failure in a tensioned shaft [5].

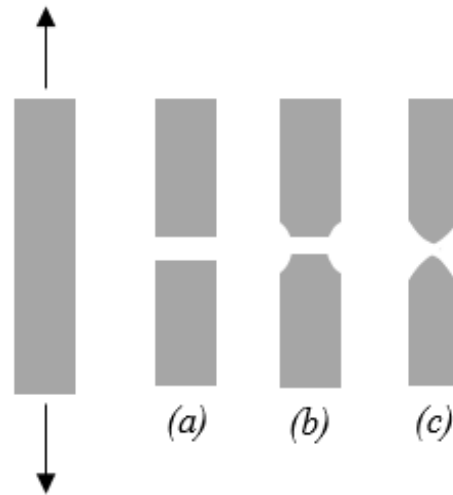


Figure 2. Failure of a shaft in tension. (a) Brittle fracture, (b) Ductile fracture and (c) completely ductile fracture (After Teja [5])

In cases where forces are torsional, a brittle shaft will shear off along the planes perpendicular to the direction where tension is maximal. In ductile material shearing occurs in plane more perpendicular to the shaft axis. Figure 3 shows the typical torsional failure of both ductile and brittle materials.



Figure 3. Failure of a shaft in torsion (a) Ductile fracture and (b) Brittle fracture (After Ahmed [6]).

In cases of compression, the diameter of a ductile shaft will increase initially as there is an elastic region and a plastic region in the response. Brittle materials have an initial linear region followed by a region where the shortening increases at a higher rate [1, 6].

Fatigue commonly occurs as a result of varying load being applied continuously over many cycles. This leads to stress in the shaft which causes micro-cracking eventually resulting in a break or fracture [1, 2, 7]. Fatigue failure usually starts with crack initiation, followed by crack propagation and a final fracture, as illustrated in figure 4.

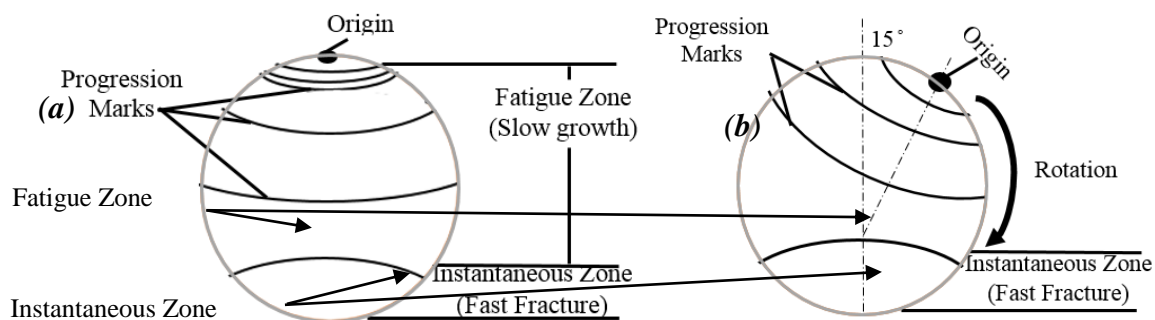


Figure 4. Progression of fatigue cracks in a shaft (Cross section. After Sachs [1])

In figure (4a), it can be seen that under bending a fatigue crack is initiated perpendicular to the origin and over time it spreads across the fatigue zone accompanied by little plastic deformation. When it reaches the edge of the so called “instantaneous zone”, the growth rate of the crack increases dramatically and complete fracture results. Figure (4b) illustrates a typical example of rotating shaft in torsional failure. The fracture begins at a single origin, and cracks develop at an angle at numerous locations around the shaft. If the

fatigue load is constant, the shaft will not leave progression marks and if the shaft is heavily loaded, the instantaneous zone will be large [1].

The extensive use of maintenance philosophies/strategies has proved to be a valuable technique to detect common failures. Such strategies aim to facilitate longer use of the machine part, improve process operation and reduce costs. These strategies include: preventive, predictive and proactive maintenance. Traditionally, a reactive “Run-To-Fail” approach is used where the equipment failure process no significant risk and the cost of other strategies is prohibitive. Thereafter, a proactive strategy is normally adopted to achieve improved functionality of the maintained part and reduce the rate of deterioration by scheduling a maintenance operation. The root cause of failure may be recognised and action can be taken to avoid machinery downtime [8, 9].

Many condition monitoring techniques are presently used to diagnose the condition of mechanical components, for example: vibration analysis, acoustic emission analysis, oil debris analysis and image processing techniques [10]. Vibration analysis, acoustic emission analysis and oil debris analysis can provide early information prior to complete failure in machine elements such as gears and bearings [11, 12, 13, 14].

Vibration analysis is commonly used to monitor machinery condition. It provides information about the defective element as well as offering modestly early indication of failure. Vibration data has been used in time, frequency, time-frequency and order domain analyses. The approach used depends on the type of application and the required information. For instance, a time domain approach has been used to analyse vibration data to predict gear fatigue failures [11] and to identify abnormal conditions in automotive mechanical transmissions [13].

Acoustic Emission has been used to monitor the sound waves that propagate through the surface of the material in the frequency range of 20 kHz – 1 MHz in machine elements and several publications have successfully addressed the use of acoustic emission to monitor failures in components such as vehicle wheel bearings [13] and fatigue gearbox splines [15].

This paper describes an online monitoring system which is capable of monitoring fatigue failure in rotating shafts. It is proposed that, the use of several-sensors and an embedded controller can form an affective failure detection system. The paper presents the details of a shaft fatigue failure detection system and describes a test bench for the purpose of capturing common mechanical failures related to rotorcraft.

2. EXPERIMENTAL TEST BENCH

A test bench was developed to capture different types of failure that occur in different elements such as the main rotary shaft and bearing. The test bench was constructed using a metal base incorporating grooved aluminium extrusions which ease the fastening of other components. The test bench is illustrated in figure 5.

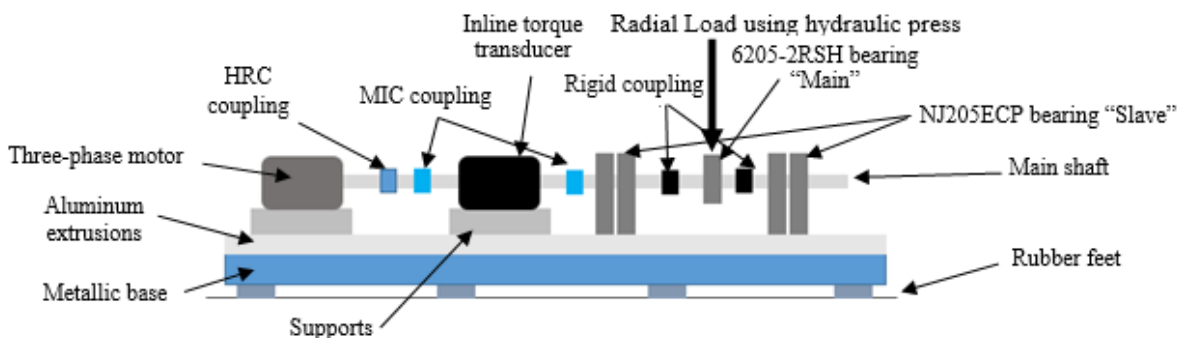


Figure 5. Schematic of test apparatus

A number of components were incorporated in the design of the hardware, these included a three-phase motor, main shaft support, bearings and torque meter. The main rotary shaft is driven by a 1.1kW variable

speed electric motor controlled by an Optidrive E² inverter capable of providing operating speeds from 0 – 3,000 rpm (0 – 50Hz). (The electric motor has a maximum speed of 2,800 rpm [16, 17].)

The motor shaft is linked to an inline torque meter (TM309) with a nominal rated torque of 20Nm. It uses a non-contact differential transformer to measure bearing torque and is connected to the main test shaft using a coupling. Specifications for the torque meter are given in table 1.

Table 1: Torque meter specifications

| Transducer model | Nominal rated torque | Maximum speed | Operating voltage |
|------------------|----------------------|---------------|-------------------|
| TM309 | 20Nm | 10,000 RPM | 20 to 32 VDC |

For the present study, five SKF bearings were used. These were: Four single row cylindrical roller bearings (NJ205 ECP) and a single deep-groove ball bearing (6205-2RSH) with an inner race diameter of 25 and outer race diameter of 52mm. The NJ205ECP bearings were used as “slave bearings” to support the load applied to the (lower capacity) test bearing. Each slave bearing has a dynamic load capacity of 30kN. The shaft is loaded through a rolling element bearing on the main shaft (SKF 6205-2RSH) which can, support around 14.8kN [19, 20]. The test shaft used in the study was manufactured from mild steel with around 0.15 to 0.40% of carbon content. It had a diameter of 25mm and a length of 100mm and was loaded radially using a hydraulic bench press.

Three main controllers were used in the system, these were: an NI USB6003, a DSP7000 and a cDAQ9191. The NI USB6003 is a 16-bit multifunction I/O device. It has eight analogue input terminals / channels with 100kS/s resolution and two analogue outputs. The Digital Signal Processing (DSP7000) is manufactured by Magtrol and is designed to be used with the inline torque transducer (TM309). It is a high speed programmable controller with the ability to read 500 torque and speed samples per second. The cDAQ9191 is 24-bit resolution controller with the ability to measure data using four accelerometer channels at a rate of 51.2kS/s.

These controllers were interfaced to sensors, these were for temperature measurement and pressure measurement in the hydraulic loading system (to assess the load on the bearing/shaft), shaft speed measurement, bearing system torque acquirement and vibration analysis [21, 22].

3. ONLINE ACTIVE HEALTH MONITORING

An online active health monitoring system was developed to identify fatigue failure in the main rotary shaft by monitoring the variation in the vibration signal using a PCB ICP uniaxial piezoelectric transducer (sensitivity 100mVg⁻¹, resolution 0.002 g rms, frequency range 2.5 to 10 kHz) coupled to cDAQ9191 controller.

Vibration data was parameterised in three domains. These were the time-domain, frequency-domain and time-frequency domain. In the time-domain three main parameters were evaluated: the RMS level, the Crest Factor and signal Kurtosis. These parameters were used to identify the level of energy in the test component as well as to measure the “peakedness” of the time series signal (TSS) [12].

The RMS value was determined from:

$$RMS = \sqrt{\frac{1}{N} \left[\sum_{i=1}^N (x_i)^2 \right]} \quad (Eq. 1)$$

Where x_i = the time signal sampled, N = the number of samples and i = the sample index.

Crest Factor (CF) is defined by:

$$Crest\ Factor\ (CF) = \frac{Peak\ value}{RMS} \quad (Eq. 2)$$

and kurtosis (Ku) is defined by:

$$Ku = \frac{N \sum_{i=1}^N (x_i - \bar{x})^4}{[\sum_{i=1}^N (x_i - \bar{x})^2]^2} \quad (\text{Eq. 3})$$

A Frequency domain technique was also utilised to help identify defects in the test element. Narrowband spectra were evaluated using a Fast Fourier Transform (FFT) tool, with the aim of detecting changes in amplitude of specific components as a sign of developing failure.

The time–frequency domain was employed to provide a joint distribution information to find the frequency of the any signal component changes as a function of time to identify if changes were visible in both the time and frequency domain.

In addition to vibration, the system also directly monitored a number of other parameters including: temperature, radial load, speed and torque, as depicted in figure 6.

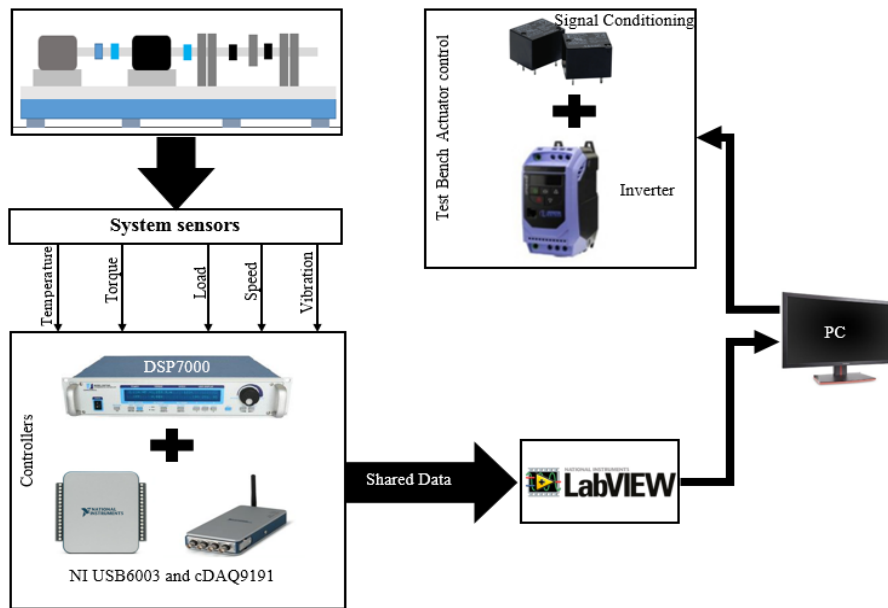


Figure 6. Main components of the active online health monitoring system

Temperature was measured by using the NTC output and the Steinhart-Hart equation. The NTC measures the temperature by sensing the variation of resistance (sensitivity $10\text{k}\Omega/25\text{C}^\circ$) and using the acquired resistance in the following equation.

$$\frac{1}{T (K)} = A + B * \ln(R) + C * \ln(R^3) \quad (\text{Eq. 4})$$

Where A, B and C are Steinhart coefficients and R is the measured resistance in ohms [23]

The radial load was monitored using a pressure sensor to measure the oil pressure in the hydraulic jack. This transducer was interfaced with an embedded system that measures the amount of pressure applied and using Pascal's law calculates the applied load. The specifications of the pressure sensor are given in table 2.

Table 2: Pressure transducer specifications [24]

| Input | Output | Working temperature | Weight |
|------------|---------------------------------|---------------------|--------|
| 0-2500 psi | 0.5 – 4.5 V DC (Linear Voltage) | -40C to + 120 C | 0.1Kg |

The online active health monitoring system measured friction torque using a TM309 torque transducer. The measuring system in the transducer has two concentric cylinders, a deformation zone, and two concentric coils (primary and secondary) attached to the housing. Both cylinders have a coincident row of slots that rotate with the shaft inside the coil. An alternating current with a frequency of 20kHz flows through the

primary coil, if torque is applied, the slots begin to overlap in the deformation zone. This gives a voltage proportional to torque in the secondary coil and the conditioning electronic circuit inside the sensor converts the voltage to a nominal torque signal varying from 0 to $\pm 5V$.

The speed was measured in the TM309 using a built-in optical sensor. It works by means of an inductive probe. It read the speed using a toothed disc on the measuring system inside the TM309 and, the output of electronic conditioner was a frequency signal proportional to the shaft rotational speed [26].

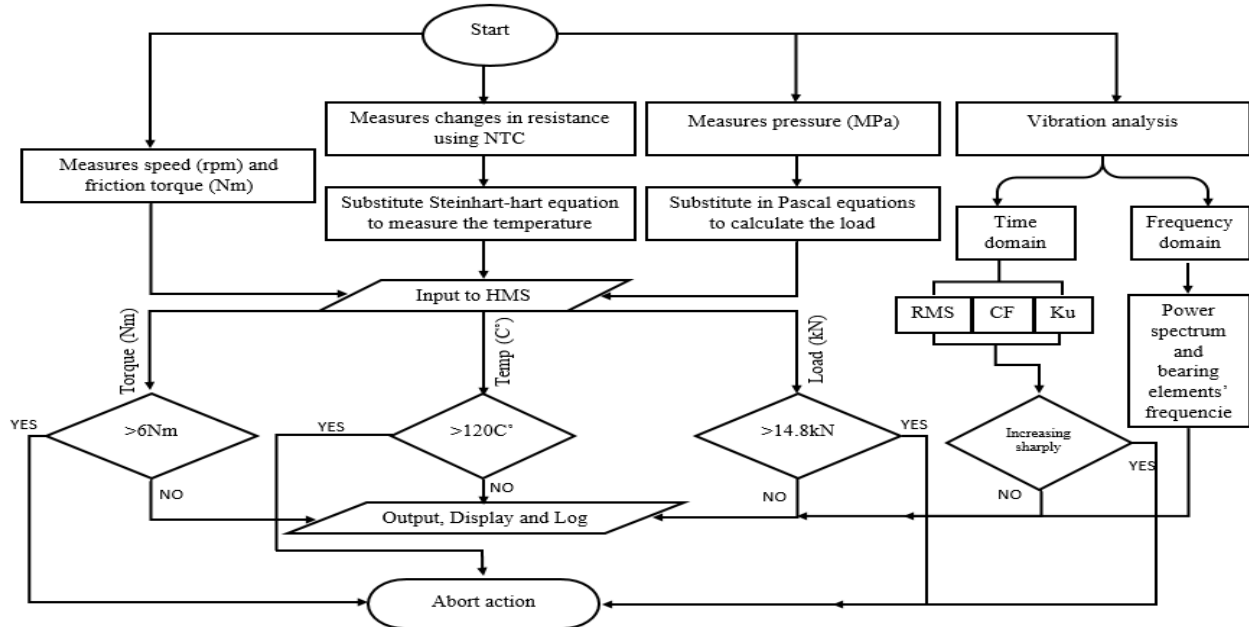


Figure 7. Software development flowchart

Code was also developed in LabVIEW with the aim of capturing the progression of fatigue. This code was able to subtract two acquired images using IMAQ tools in LabVIEW. Firstly, it created an image buffer for two acquired images, the initial image and second image from a later stage of operation (Images A and B). Then these images were loaded into a buffer, where they were converted to grayscale. The user uploads these images, allowing a region of interest (ROI) to be cropped and allowing the two images to be registered and subtracted by visual inspection. The difference map generated contains features that illustrate progression of the fault (fracture).

4. TESTING AND SYSTEM VALIDATION

The experimental setup was used to collect the signals from an online health monitoring system during a fatigue failure process. The shaft was loaded radially using a four tonne hydraulic bench press. Three NTC temperature sensors were incorporated into the system and mounted in different places: under the main bearing to measure the operating temperature and attached to the hydraulic press piston to measure transferred heat, as well as being located in the test bench to measure the cabinet temperature. All tests were carried out at 2000 rpm with 10kN of radial load applied to the shaft. Operationally, 10kN was applied to the shaft initially then the NI USB6003 controller requested 2000 rpm as a demand speed incrementally.

A Cannon EOS 750D was utilised to capture images in this experiment. It recorded the rotational movement of the main shaft, to allow information about the effect of fatigue progression and machine performance to be extracted from images. Using the video recorded during experiments, LabVIEW code was used to identify the fatigue progression. Graphs in the following sections illustrate some of the data logged from sensors.

Two tests were conducted, these are described below.

Test 1: In this test the bearing was located loosely on the test shaft and held in place by two external circlips to resist the axial movement while loaded. The circlips were positioned in two 1mm deep grooves around the shaft and the shaft was loaded at 10kN while operating at 2000 rpm.

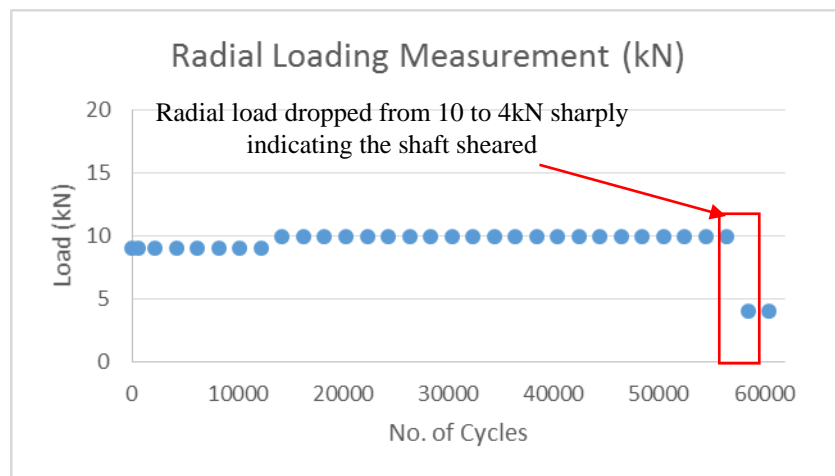


Figure 8. Applied radial load as a function of test cycles

It was noticed in this test that the applied load increased slightly with time, as illustrated in figure 8. This was unintended and it was assumed to be due to an increase in temperature inside the test cell due to the operation of the motor and dissipation of frictional heating from the bearing which subsequently caused expansion of the oil inside the hydraulic cylinder leading to increased cylinder pressure and thus increased load.

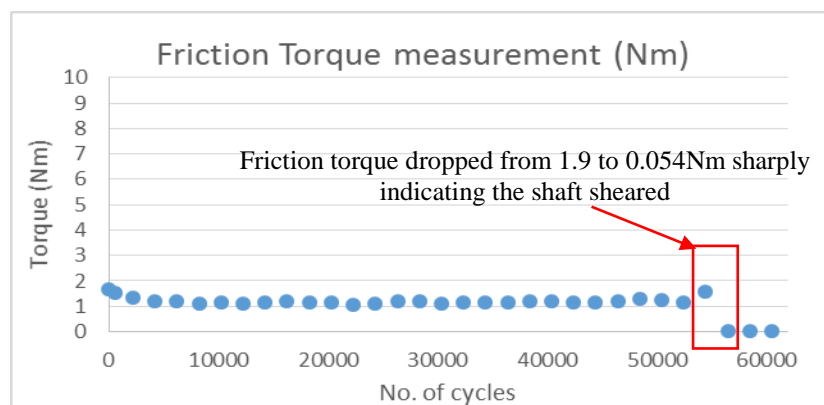


Figure 9. Friction torque as a function of test cycles

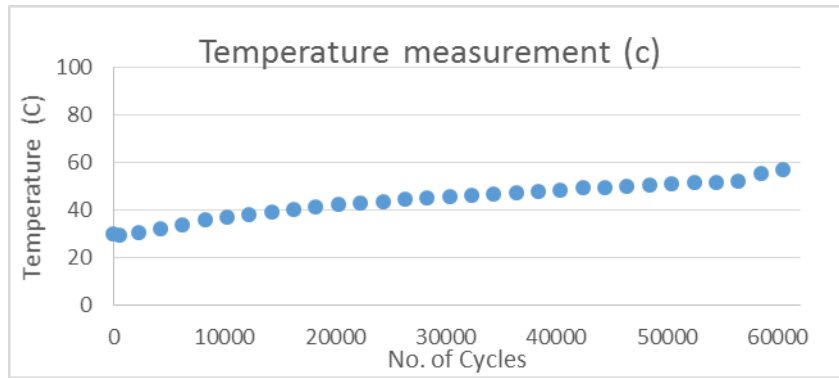


Figure 10. Operating temperature of the shaft.

The later stages of the graphs in figure 8 and figure 9 show a sudden drop in the load and torque, these signals identify a developing failure and offer the opportunity to be used as part of a set of criteria for the initiation of an automatic shutdown action. The change in the temperature of the shaft is shown in figure 10 and the fully sheared shaft is illustrated in figure 11.

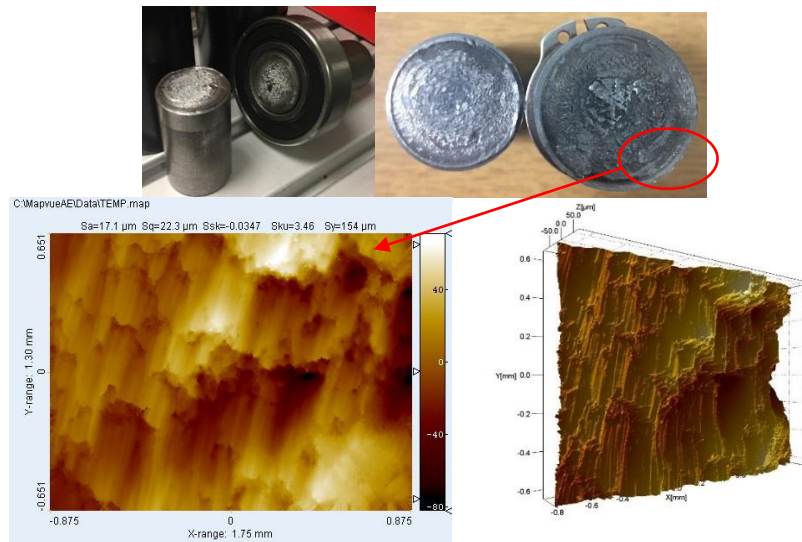


Figure 11. Macroscopic and fractography images of the sheared shaft.

Test 2: The second test did not involve any external components attached to the shaft such as circlips and had no features such as grooves which may promote stress concentration leading to early onset of failure. The sampling rate was changed to obtain a greater degree of in depth information on the parameters recorded in this test. The following graphs illustrate the results that were obtained.

Figure 12 presents a record of the the applied radial load during the last phase of the test up to the point where the shaft shears identified by a fall in friction torque.

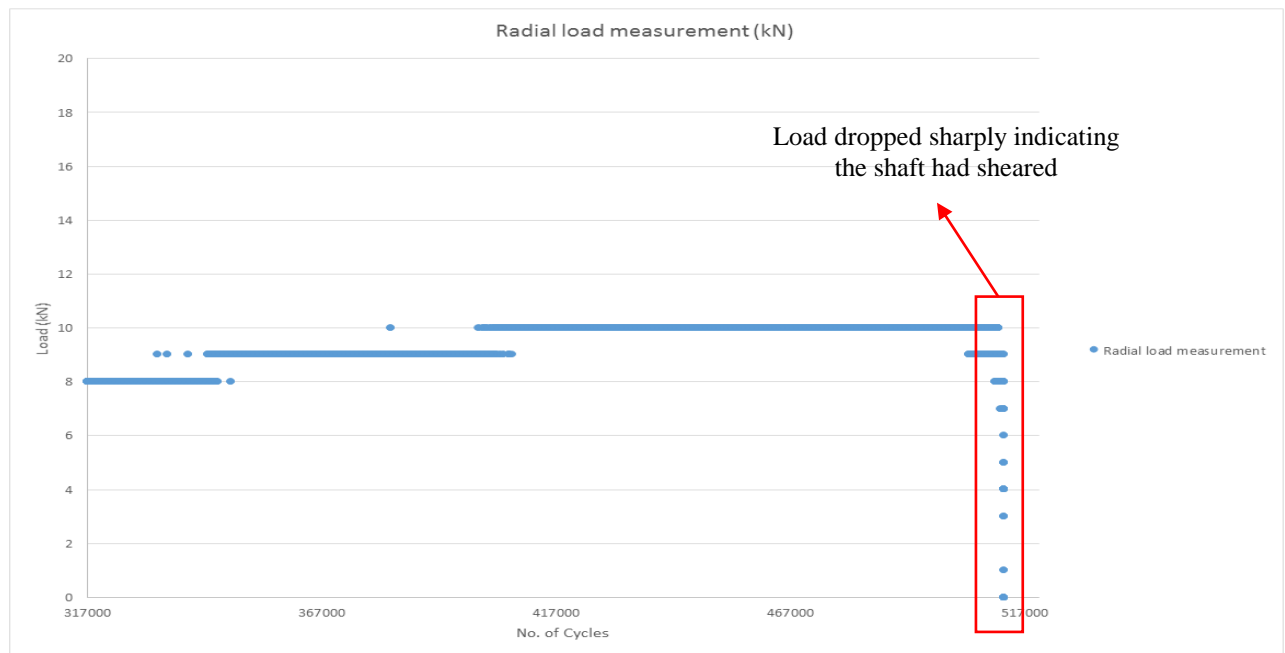


Figure 12. Shaft load in test 2

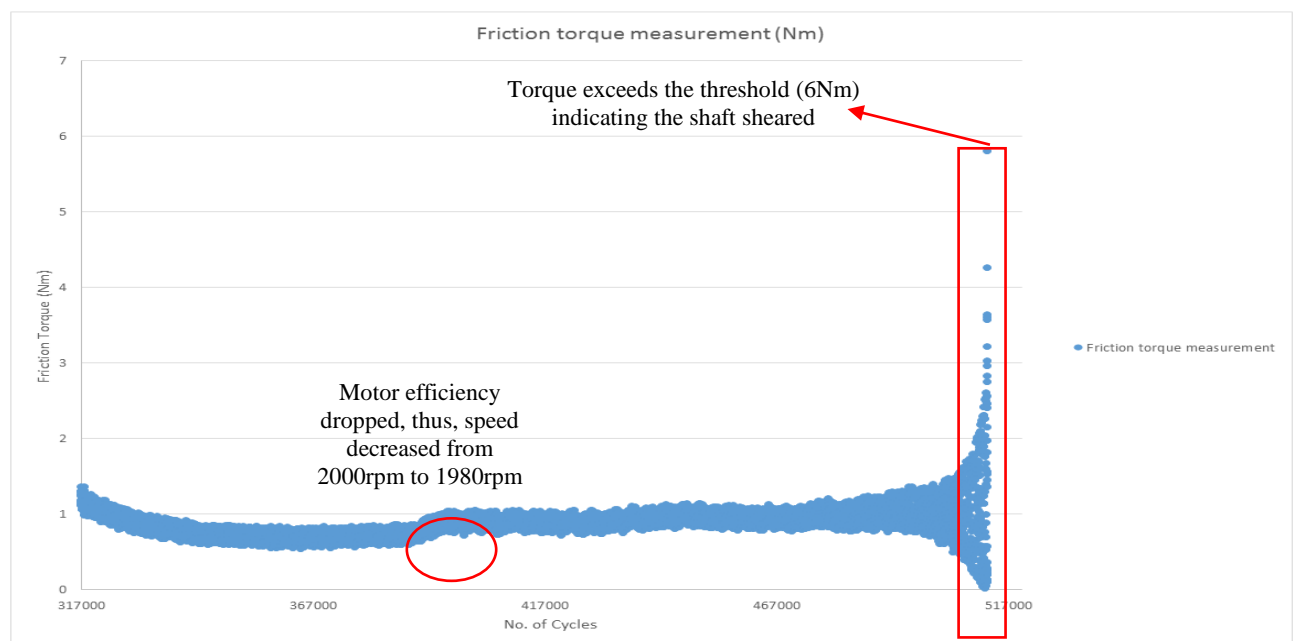


Figure 13. Friction torque in test 2

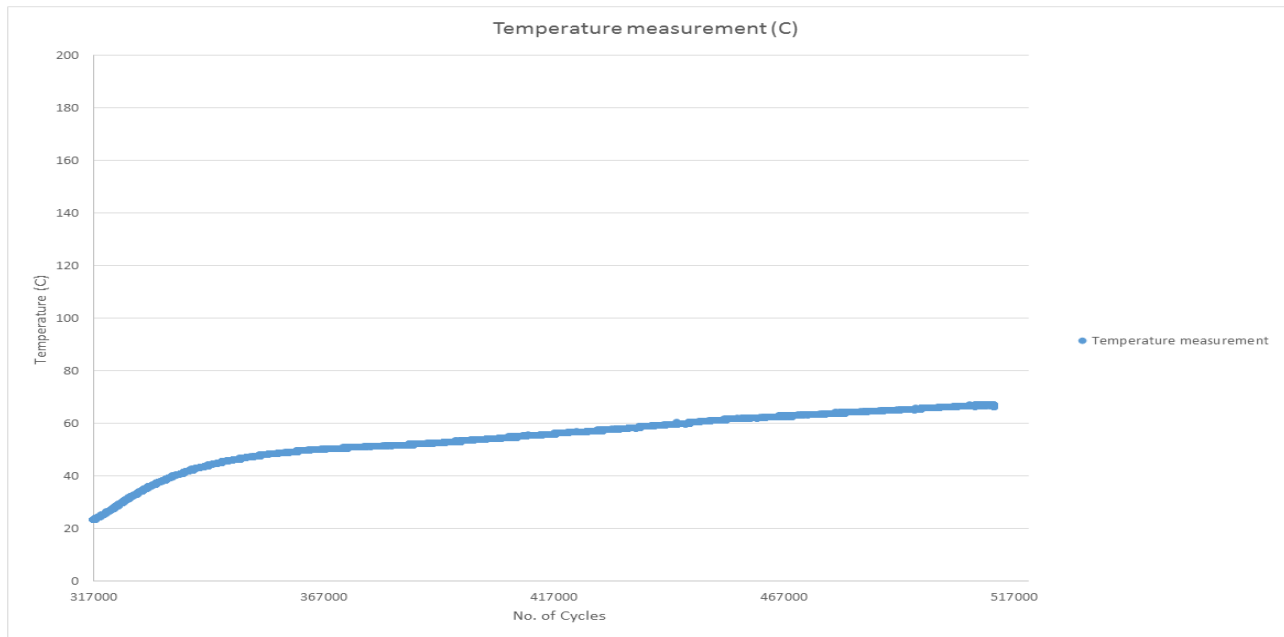


Figure 14. Operational temperature of shaft (C)

In figure 13 it can be seen that as the shaft approaches failure the bearing torque becomes increasingly erratic and becomes very large as shaft failure is reached. Figure 14 illustrates the temperature of a bearing holder during the final stage of the test. For most of the test the hydraulic bench press piston acted directly on main bearing. It was found that the bearing heated linearly at a modestly high rate under these conditions. This arrangement was replaced by a bearing holder, separated from the hydraulic piston by a ceramic insert to reduce the conduction of heat between the two components at an interim stage in the test. The development of temperature of the holder was then found to follow a different pattern as the holder had more heat capacity than the bearing alone. Figure 15 presents images of the fully fatigued component, the various stages of the fatigue process are visible in the image. Figure 16 details changes in the derived vibration parameters during the test. Figure 17 illustrates a sample frequency spectrum taken approximately one minute before complete failure of the shaft. Figure 18 presents the intensity of the spectral harmonics vary over a short one second interval at approximately the same stage of the test as figure 17.



Figure 15. Fatigue surface of the second sample – (i) fatigue initiation zone, (ii) fatigue zone (slow crack growth), (iii) instantaneous failure zone (rapid failure)

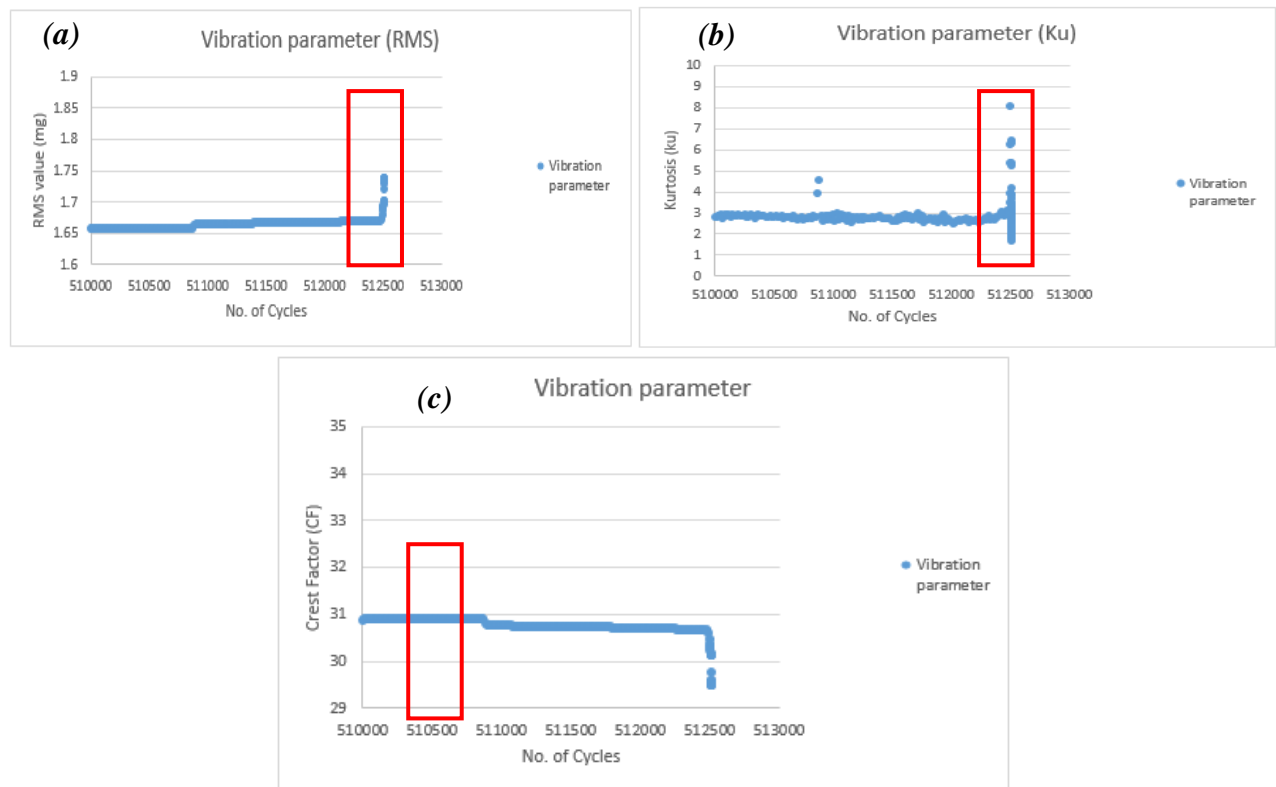


Figure 16. Vibration parameters from final stage of test including: RMS, Kurtosis (Ku) and Crest Factor (CF).

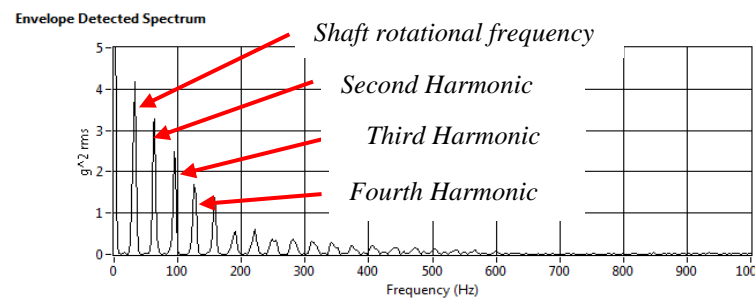


Figure 17. Frequency-domain spectrum of shaft vibration.

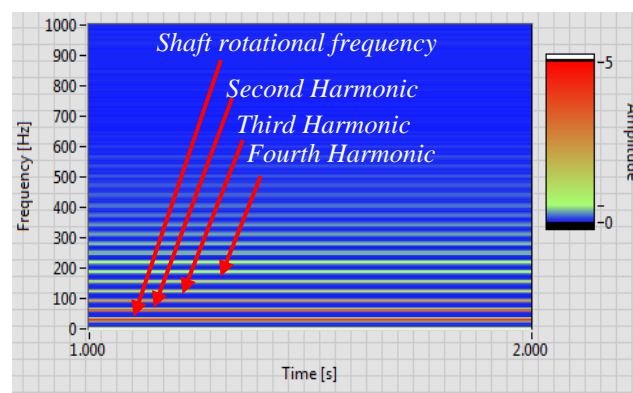


Figure 18. Time – Frequency domain technique illustrating the shaft frequency and its harmonics.

11

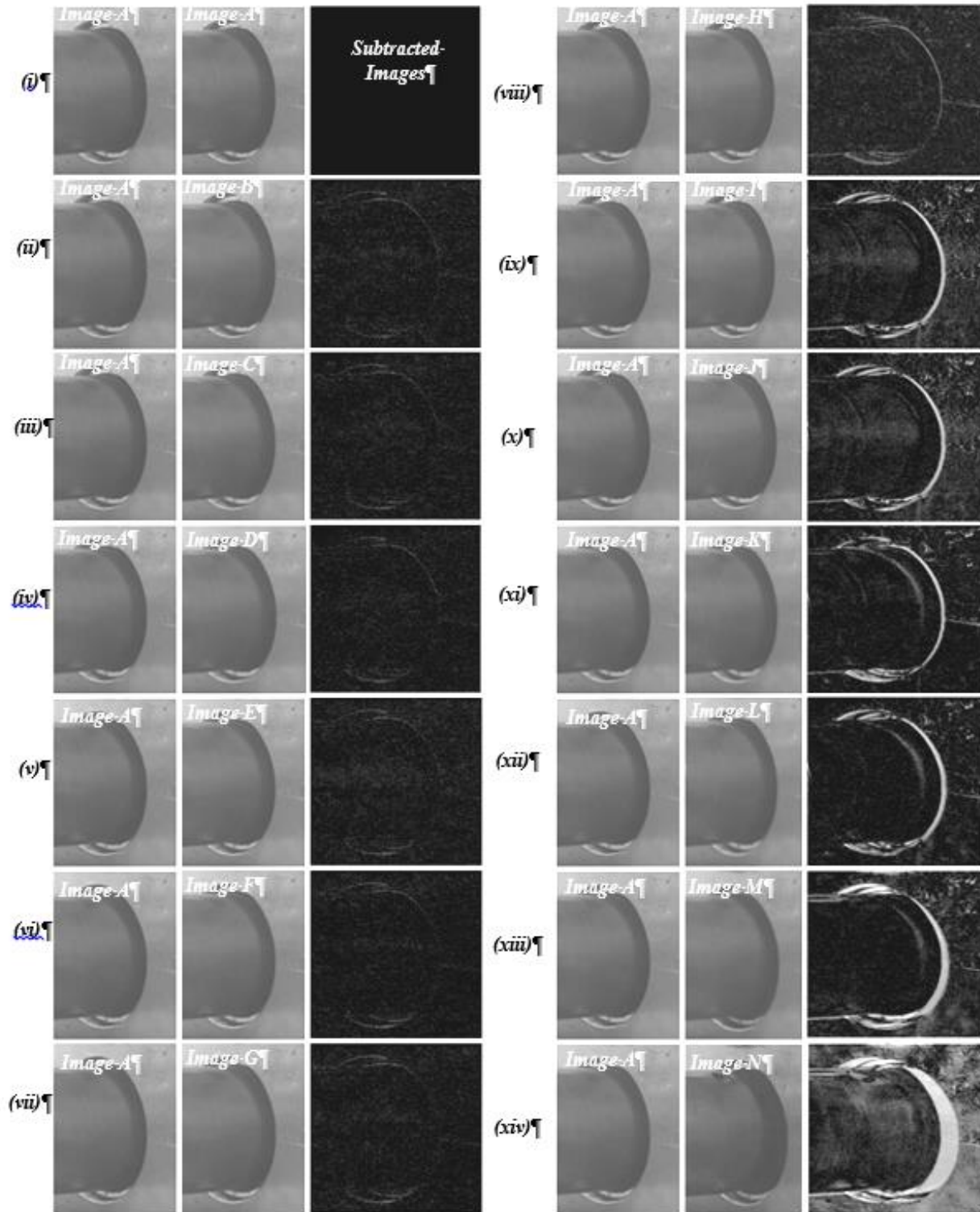


Figure 19. Images from surveillance video – (Image A) captured at the outset. Image (B) captured after the first 50 second, Image (C) after 1 minute, Image (D) 1m 30sec, Image (E) 2 minutes, Image (F) 2 min 30 sec, Image (G) 2 min 40 sec, Image (H) 2 min 50 sec, Image (I) 3 minutes, Image (J) 3 min 30 sec, Image (K) 4 minutes, Image (L) 6 minutes, Image (M) 9 minutes and Image (N) at 9 min 17 sec.

To facilitate detection of failures using image processing, the experiment was monitored by a stationary camera mounted on a stand. Video images were acquired during the experiment and sample frames from the last 10 minutes of the life of the shaft are illustrated in figure 19. Images like this may prove to be useful source of information in future failure analysis as they can identify developing features related to displacement due to failure. Image (A) represents a reference image taken at the beginning of the experiment and Image N represents the point at which the shaft was fully sheared. Images of the shaft and

bearing holder were subtracted from the reference, image A, at different times during the experiment. It can be seen from the difference images in figures 19 (viii – xiv) that as the failure develops, motion of the parts due to the reduced stiffness arising from crack development, combined with the effect of vibration, gives a generally increasingly defined outline of the separation between system elements which is progressive. Inspection of the difference images (ii – vii) reveals that during normal operation there is a slight difference in the position of some components. In figures 19 (viii, ix, x, xi, xii and xiii) these differences become more evident, indicating that the bearing holder is further displaced as a consequence of the increased freedom given due to a developing fatigue crack. Finally, case (xiv) shows that complete shaft fracture has entirely removed part of the positional constraint for the bearing holder position and led to significant differences in parts of the image as revealed by the difference analysis. Such feature changes could, of course, be enhanced and quantified in significant detail using sophisticated image analysis, but such analysis has not been done here.

5. FURTHER DEVELOPMENT

The data collected in this investigation will be used to inform development of an active condition monitoring system. It is planned that active health monitoring will be based on a thresholding algorithm. Each relevant parameter being ascribed a threshold, if the limit of threshold conditions is exceeded, automatic action will be taken. In the first instance, such action would consist of shutting the whole system down to prevent damage to the other elements within the application. A range of monitoring strategies will be included. For example, according to the specification of the “Test” bearing, it can resist a maximum dynamic load of 14.8kN and maximum operating temperature is 120C°, thus, either of these limits is exceeded, the system will be closed down. Additionally, a fatigued shaft displays a higher friction torque in its bearings than an undamaged shaft [27, 28]. This is due to varying misalignment in the support bearings, this effect also causes vibration. Torque was initially measured while the system was loaded, where a clear limit was evident. In the final design, if the torque exceeds a specific limit, the system will shut down. Additionally, if the torque value falls and the motor speed remains at the same level, the system will also be closed down as the shaft will have become disconnected.

The system will be aborted using a low-cost mechanical relay (JZC-11F). It uses a Single Pole Double Throw (SPDT) configuration that requires 5V DC to switch between the throw. It has a maximum switching current of 5A and a maximum switching voltage of 30V DC / 25V AC [26]. When the thresholding algorithm detects that any of the conditions explained above are met, a digital signal will be sent from USB6003 to the relay board to switch between the terminals in the inverter resulting in the motor shutting down.

6. CONCLUSION

This paper presents the details of a test bench, developed by the authors, to study methods for online health monitoring of bearings and shafts. In a final state, it is proposed that an automatic control system will incorporate a series of electronic sensors and controllers in conjunction with a simple algorithm control to form an active online health monitoring system. The sensors will be supported by video surveillance of the parts to allow simple image processing to be used to track failure progression.

REFERENCES

- [1] Sachs, N. (2012). Failure Analysis of Machine Shafts, [Online] Maintenance Technology, Available at: <http://www.maintenancetechnology.com/2012/07/failure-analysis-of-machine-shafts/> [Accessed 25 Mar. 2017].
- [2] Li, W., Xia, F. and Luo, B. (2016). Failure Analysis of a High-Speed Shaft Crack. *Journal of Failure Analysis and Prevention*, 16(5), pp.919-930. doi: 10.1007/s11668-016-0172-4
- [3] Sumit P. R., Laukik P. R (2014). A Review of Various Techniques Used for Shaft Failure Analysis, *International Journal of Engineering Research and General Science*, 2(2), pp: 159 – 171, ISSN 2091-2730.
- [4] Guimaraes, A., Brasileiro, P., Giovanni, G., Costa, L. and Araujo, L. (2014). Failure analysis of a half-shaft of a formula SAE racing car, *Case Studies in Engineering Failure Analysis*, vol. 7, pp: 17-23, doi: 10.1016/j.csefa.2016.05.002.
- [5] Teja G. V., V. (2014). Chapter IV – learning from failures. [Online] Slideshare.net. Available at: <https://www.slideshare.net/varunteja7330/mi-291-chapter-4-learning-from-failure> [Accessed 25 Mar. 2017].

- [6] Ahmed, T. (2013). Torsion force. Slideshare.net, doi: 10.01.03.069.
- [7] Rossino, L., Castro, D., Moreto, J., Ruchert, C., Spinelli, D. and Tarpani, J. (2014). Surface contact fatigue failure of a case hardened pinion shaft, vol.17, n.3, pp.535-541, ISSN 1516-1439, doi:10.1590/S1516-14392014005000063
- [8] Lura S. (2001) Linking maintenance strategies to performance, International Journal of Production Economics, 70(3), pp: 237-244, doi: 10.1016/S0925-5273(00)00067-0
- [9] Smoczek, J. and Szpytko, J. (2012) 'Supervisory system for supporting decision-making process in Proactive maintenance of technical object', IFAC Proceedings Volumes, 45(31), pp. 30–35, doi: 10.3182/20121122-2-es-4026.00023
- [10] Jaber, A.A. and Bicker, R. (2014) 'The State of the Art in Research into the Condition Monitoring of Industrial Machinery', International Journal of Current Engineering and Technology, 4(No.3), E-ISSN 2277 – 4106, P-ISSN 2347 - 5161.
- [11] Onsy, A., Bicker, R. & Shaw, B.A. (2010) Intelligent Diagnostic Health Management of Power Transmission Systems: An Experimental Validation, International Journal of COMADEM 2010, 13(2), 46-58. ISSN 1363-7681
- [12] Onsy, A., Shaw, B.A. & Jishan, Z. (2011) Monitoring the Progression of Micro-Pitting In Spur Geared Transmission Systems Using Online Health Monitoring Techniques. SAE International Journal Aerospace, 4 (2), doi: 10.4271/2011-01-2700.
- [13] Onsy, A. (2013) A New Acoustic Emission Remote Sensing System: an Experimental Validation of Wheel Bearing Condition Monitoring. SAE International Journal Aerospace, 6(2):2013, doi: 10.4271/2013-01-2221.
- [14] Helmy, M., Onsy, A., Hussein, W. M., & El Sherif, I. (2014) Development of an Advanced Diagnostic System for Automotive Mechanical Transmissions. International Journal of COMADEM. 2014, 17(2), 39- 44. ISSN 1363-7681
- [15] Zhang, L., Ozevin, D., Hardman, W., Kessler, S. and Timmons, A. (2016). Fatigue crack growth monitoring of idealized gearbox spline component using acoustic emission. Non-destructive Characterization and Monitoring of Advanced Materials, Aerospace, and Civil Infrastructure, Volume 9804, ISBN: 9781510600454, doi: 10.1117/12.2220163.
- [16] TEC Electric Motors Ltd. (2010) TEC Electric Motors Catalogue V3: User manual. Retrieved from: <https://inverterdrive.com/file/TEC-AC-Motor-Technical-Catalogue-IE1-IE2-v3> [Accessed 25 Mar. 2017]
- [17] Invertek Drives (2009) Optidrive E2 single phase output: User manual, Retrieved from: <http://www.invertekdrives.com/variable-speed-drives/optidrive-e2-single-phase> [Accessed 25 Mar. 2017]
- [18] Magtrol (2015) Series Miniature Inline Couplings: User manual, Retrieved from: <http://www.magtrol.com/manuals/micmanual.pdf> [Accessed 25 Mar. 2017]
- [19] SKF (2017) Single row Deep-groove ball bearing – 6205-2RSH: User manual, Retrieved from: <http://www.skf.com/uk/products> [Accessed 25Mar. 2017]
- [20] SKF (2017) Single row Cylindrical roller bearings – NJ 205 ECP: User manual, Retrieved from: <http://www.skf.com/uk/products> [Accessed 25Mar. 2017]
- [21] Magtrol (2015) DSP7000 series high-speed programmable Dynamometer controllers: User manual, Retrieved from: <http://www.magtrol.com/manuals/dsp7000manual.pdf> [Accessed 25 Mar. 2017], pp: 2 – 10.
- [22] National Instrument - NI (2014) 6003 device specifications: User manual, Retrieved from: <http://www.ni.com/pdf/manuals/374259a.pdf> [Accessed 25 Mar. 2017]
- [23] Stanford research systems (2012) Calibrate Steinhart-Hart coefficients for Thermistors, [Online], Available at: <http://www.thinksrs.com/downloads/PDFs/ApplicationNotes/LDC%20Note%204%20NTC%20Calculator.pdf> [Accessed 25 Mar. 2017]
- [24] CPC (2000) CPC Pressure Transducer Specifications and Installation Instructions: User manual / guide, Retrieved from: <http://www.cpcus.com/> [Accessed 25 Mar. 2017]
- [25] Magtrol (2013) Series in-line torque transducers: User manual, Retrieved from: <http://www.magtrol.com/datasheets/tm301-308.pdf> [Accessed from 25 Mar. 2017]
- [26] Peng, C., Xianke, H. and Jianming, X. (2009a) Miniature Relay – JZC-11: User manual, Retrieved from: <https://www.sparkfun.com> [Accessed 25 Mar. 2017]
- [27] R. A. Gujar, S. V. Bhaskar (2013) Shaft Design under Fatigue Loading By Using Modified Goodman Method., International Journal of Engineering Research and Application (IJERA), 3(4), pp: 1061-1066, ISSN: 2248-6922
- [28] A. Dindar, M. Akkök and M. Çalışkan (2016) Experimental Determination and Analytical Model of Friction Torque of a Double Row Roller Slewing Bearing, Journal of Tribology, 139(2), doi: 10.1115/1.4033364

Session 13

Advanced Signal Processing and Big Data 1

Session Chair Dr. Ming Zhao

Bearing Fault Feature Detecting Based on Nonlocal Means Denoising

(Y. Wang, S. He, S. Zhang)

The Influence of B-Value Distribution in Damage Evaluation of Structural Material Based on AE Parametric Analysis *(M. T. I. Khan, S. Tsurumaru and R. Hidaka)*

A Self-Adaptive AE Signal Classification Method on Energy Domain *(Y. Zhou, D. Wang, L. Lin)*

Bearing Fault Feature Detecting Based on Nonlocal Means Denoising

Yanxue Wang, Shuiliong He and Suofeng Zhang

School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, PR. China

ABSTRACT

Defective rolling bearing response is often characterized by the presence of periodic impulses, which are usually immersed in stochastic noise. In order to more effectively identify incipient defect, nonlocal mean denoising is applied to extract features in this work. The performances of nonlocal mean denoising are evaluated in comparison with wavelet threshold denoising, translation invariant wavelet denoising and the second generation wavelet denoising. Vibration signals acquired from bearings with defect in the outer raceway, inner raceway and rolling element fault, respectively, extracted availably based on the proposed approach. Experimental results show that the proposed approach is feasible and effective to detect roller bearing faults from vibration signals.

Keywords: Nonlocal means denoising, feature detection, denosing, rolling bearing, fault diagnosis.

Corresponding author: Y. Wang (email: wyx1999140@126.com)

1. INTRODUCTION

Roller bearings are commonly applied to various rotating machines, and play a significant role in the modern manufacturing industries [1-2]. Therefore, fault detection of roller bearings is of great importance in order to avoid accidents caused by defects of bearing. Currently, vibration signal analysis is adopted to detect bearing defects [3]. However, bearing features induced by defects are often mingled with random noise in practice. Consequently, many signal denoising techniques have been proposed to improve the performance of bearing fault detection. The traditional noise reduction method is based on the spectral characteristics of the signal, while the random noise in the observed signal is depressed by designated filters. Nevertheless, this kind of method has the contradiction between the protection of local characteristic of the signal and the noise suppression, which are not well satisfied with the requirements of the vibration signal denoising. Wavelet transform (WT) has an advantage of time-frequency localization characteristic, thus it is extensively used to analyze the non-stationary signals. Signal denoising with threshold in wavelet domain have achieved good results in the field of signal processing [4, 5], but they often suffer from the pseudo Gibbs phenomena in the process of threshold denoising. Translation invariant wavelet (TIW) method is a great improvement of traditional wavelet threshold, which can effectively reduce the random noise as well as the pseudo Gibbs phenomenon [6]. Moreover, the aforementioned wavelet threshold denoising methods also have the disadvantage on selecting the appropriate base function. Second generation wavelet (SGW) is constructed based on lifting scheme in the time domain [7], thus it has several advantages including possibility of adaptive and nonlinear design, in-place calculations, irregular samples and integral transform [8] compared with classical WT. It is worth mentioning that the performance of the aforementioned denoising techniques more or less affect those desired signal components.

Unlike some methods based on the so-called locally adaptive recovery paradigm [9], a novel nonlocal means (NLM) algorithm has been applied to denoising in biomedical signals because of its better performance in preserving sharp edges of the given signals [10]. NLM filter was initially used in image denoising which addresses the problem of edge degradation, proposed by Buades in [11]. Since then, NLM algorithm has been widely researched and extensively adopted in two-dimensional (2-D) images processing

[12, 13]. Considering the influence of parameter λ to the effects of the noise reduction, Stein unbiased risk estimator (SURE)-based selection of optimal parameters is introduced into NLM algorithm with a lower computational cost [14, 15]. The non-local principle based on the fact that similar patches can occur anywhere in the image which is contributed to denoising [14]. NLM mainly uses the weighted average to remove noises by looking for similar patches in image neighborhoods, effectively avoiding the disadvantage of point based noise-removal processing. On account of defective rolling bearing response is often characterized by the presence of periodic impacts mingled with random noise, thus it is also feasible that NLM is used to remove noise from vibration signals. In this paper, NLM is introduced into vibration signal denoising prior to Hilbert envelop spectrum analysis for the purpose of bearing fault feature detection.

The rest of the paper is organized as follows: the theory of NLM algorithm is briefly introduced in Sections 2. Performances of the NLM denoising, compared with classical wavelet threshold denoising, TIW denoising and the SGW denoising are evaluated using simulated signals in Section 3. In Section 4, the effectiveness of the proposed approach is further verified using the practical bearing vibration signals. Conclusions are drawn in section 5.

2. THEORY AND METHOD

2.1. A brief introduction of NLM

Given a set of 1-D noisy signal model: $y(t) = x(t) + n(t)$ where $x(t)$ is the true signal without noise and $n(t)$ is the additive white Gaussian noise. The NLM is mainly used to recover the original signal $x(t)$ from the noisy signal $y(t)$. Thus, the recovered signal $\hat{x}(s)$ using NLM denoising can be written below,

$$\hat{x}(s) = \frac{1}{Z(s)} \sum_{t \in N(s)} w(s, t) x(t) \quad (1)$$

in which $z(s) = \sum_{t \in N(s)} w(s, t)$ is a constant used for normalization, and $N(s)$ represents an entire search domain with the center s . The size of the weight $w(s, t)$ depends on the similarity between the local patch around samples s and t , the weight $w(s, t)$ is defined as [14]:

$$w(s, t) = \exp \left(- \frac{\sum_{\delta \in \Delta} (v(s + \delta) - v(t + \delta))^2}{2L_{\Delta}\lambda^2} \right) = \exp \left(- \frac{d^2(s, t)}{2L_{\Delta}\lambda^2} \right) \quad (2)$$

where $w(s, t)$ meets two conditions $0 \leq w(s, t) \leq 1$ and $\sum w(s, t) = 1$, and d^2 denotes the Euclidean distances between samples in the patches centered on s and t which can be also considered as the degree of similarity between different patches, while Δ represents a local patch surrounding t , L_{Δ} is the local patch centered on t and λ is a bandwidth parameter. Actually, $w(s, t)$ depends on patch similarity [10] which is helpful to preserve edges. The weight is linear to the length of $N(s)$, which is usually limited by the computation burden. Hereinafter, the performance of noise reduction will be discussed in follow vibration signals.

2.2. Parameter selecting

In NLM algorithm, P , λ , M are the main parameters and the relationship of those parameters are illustrated in figure 1. It can be seen the local objective patch surrounding s is the weighted average of the local patch surrounding t , P (so $L_{\Delta} = 2P + 1$) controls the size of the similar patches and M is a half-width of the search field $N(s)$. In theory, the larger the search field $N(s)$ is, the more similar patches are. Thus a better denoising results will be achieved with a larger $N(s)$. However, the enlarged search domain also inevitably increases computational burden [8]. Moreover, the bandwidth parameter λ plays an important role in NLM, improper value of λ has an adverse effects on denoising results [14]. In order to determine the bandwidth parameter λ used in NLM method, motivated by the Stein unbiased risk estimator (SURE) based approach in [14, 15], λ is simply determined in this work as follows

$$\lambda = 0.5\sigma \quad (3)$$

In which noise variance σ can be estimated with [17]:

$$\sigma = \frac{\text{Med}(|d_j|)}{0.6745} \quad (4)$$

where d_j denotes coefficient of the j th scale, and $\text{Med}(|d_j|)$ represents the median of the sequence d_j .

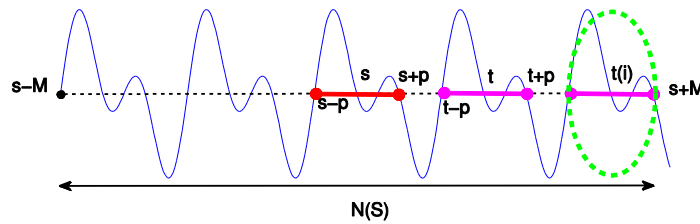


Figure 1. Illustration of NLM parameters

2.3. The proposed bearing fault feature extraction method

The proposed denoising technique combined with Hilbert envelope analysis [16] is used to detect bearing fault signatures. The process of the proposed method is schematically shown in figure 2. The effectiveness of the proposed technique will be investigated in the following section.

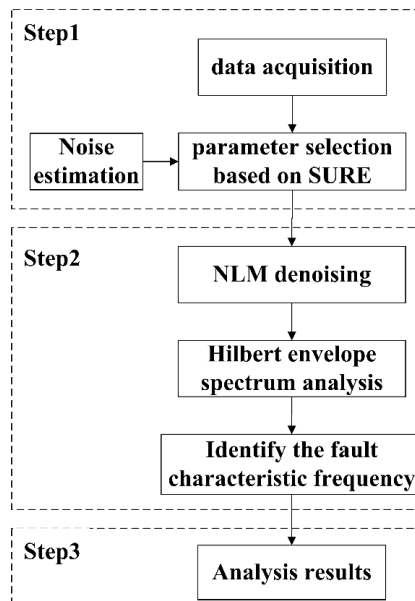


Figure 2. The flow chart of bearing fault detection approach

3. SIMULATION ANALYSIS

The performances of NLM denoising are first investigated through simulated signals in this section compared with wavelet threshold denosing (WTD) [5], translation invariant wavelet denoising (TIWD) [6] and the second generation wavelet denoising (SGWD) [8]. The simulated signal resulting from rolling bearing failure is written as follows:

$$x(t) = s(t) + n(t) = \sum_{k=0}^{\infty} A_k h(t - kT - \tau_k) + n(t) \quad (5)$$

$$A_k = 1 + A_0 \sin(2\pi f_r t) \quad (6)$$

$$h(t) = e^{-Ct} \sin(2\pi f_n t) \quad (7)$$

where $s(t)$ is the original periodic impulse signal, $n(t)$ is Gaussian white noise and T is pulse cycle. The random slippage τ_k corresponding to T , initial amplitude of the impulse A_0 , rotational frequency f_r , decay factor C and resonant frequency f_n have been fixed to 80Hz, 0, 0.3, 20Hz, 700 and 3kHz, respectively. In this paper, sampling frequency f_s is set to 12 kHz, while the data length is 2048.

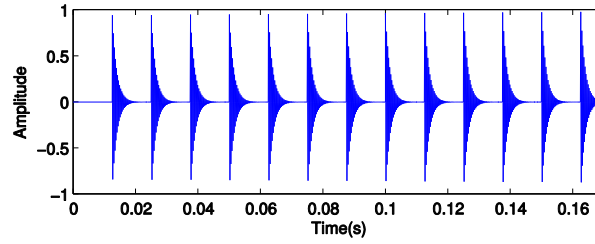


Figure 3. The flow chart of bearing fault detection approach

The simulated signal $s(t)$ without noise is shown in figure 3. The performances of the aforementioned denoising methods are evaluated based on three metrics: signal to noise ratio (SNR), root mean square error (RMSE) and percent distortion (PRD), which are defined below:

$$\text{SNR} = 10 \log \left\{ \frac{\sum_{i=1}^N s(i)^2}{\sum_{i=1}^N (s(i) - \hat{s}(i))^2} \right\} \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (s(i) - \hat{s}(i))^2} \quad (9)$$

$$\text{PDR} = 100 \sqrt{\frac{\sum_{i=1}^N (s(i) - \hat{s}(i))^2}{\sum_{i=1}^N s(i)^2}} \quad (10)$$

where $s(i)$ ($i = 1, 2, \dots, N$) is the original periodical signal, while $\hat{s}(i)$ ($i = 1, 2, \dots, N$) denotes the purified signal. As defined in Eqs (8-10), a higher SNR value and a lower RMSE and PDR values represent a better denoising result.

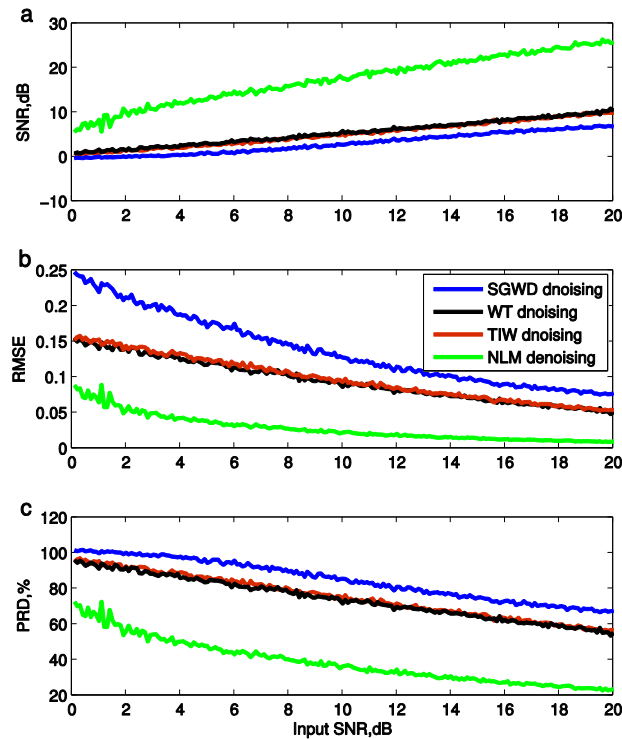


Figure 4. The comparative analysis of SNR、RMSE and PDR with de-noised signal

It can be seen in figure 4 that the NLM denoising achieves higher SNR with the lower RMSE and PDR in comparison with those aforementioned denoising methods for all the simulated signals. Therefore, the effectiveness of NLM denoising is demonstrated, then it is used as a preprocessing technique prior to Hilbert envelope spectrum analysis for vibrating signals.

4. EXPERIMENTAL EVALUATIONS

4.1. Test rig

In this section, three kinds of bearing (out-race defect, inner race defect and rolling element fault) signals acquired on MFS-Magnum test-rig are conducted to validate the effectiveness of the proposed method. The test-rig is illustrated in figure 5. Vibration data collected using a VQ data acquisition system with accelerometer (sensitivity 98 mV/g) fixed near the bearing bases. Type of ER-12K bearings are employed in experiments. Parameters relating to the fault signature are listed in table 1.

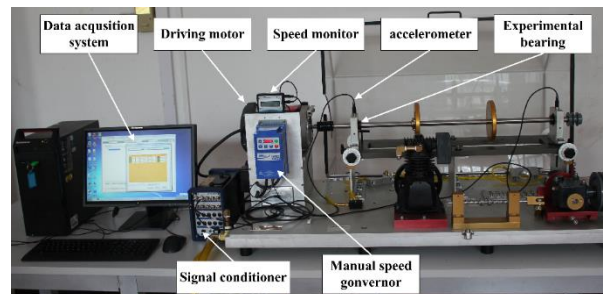


Figure 5. The machinery fault experimental platform (MFS-MG)

Table 1: Parameters of the bearing mentioned in the three cases

| Defect location | Rotational speed (r/min) | Sampling frequency f_s (kHz) | Rotation frequency f_r (Hz) | Fault characteristic frequency (Hz) |
|-----------------|--------------------------|--------------------------------|-------------------------------|-------------------------------------|
| Outer race | 1790 | 12.8 | 29.83 | 90.96 |
| Inner race | 1792 | 25.6 | 29.87 | 147.84 |
| Ball | 2390 | 12.8 | 39.83 | 79.36 |

4.2. Detecting bearing outer race fault

The raw vibration signal of bearing with an outer race defect is shown in figure 6. As is well known, vibration signal should present the impulsive features due to the presence of defect in the rolling bearing. However, weak signatures in the time domain are often contaminated because of the existing strong ambient noises.

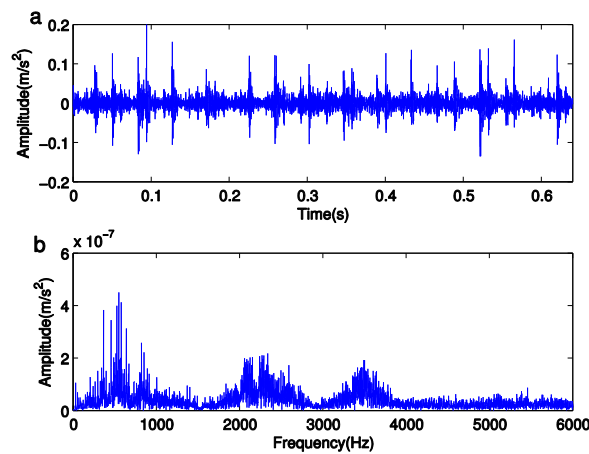


Figure 6. Vibration signal of bearing with an outer race fault time domain waveform (b) FFT spectrum

To eliminate the noise and enhance the fault information, the aforementioned denoising method is applied to process the bearing vibration signal, respectively. The time domain result of the purified signal are shown in figure 7. It can be seen that NLM algorithm can efficiently eliminate noise, compared with three other denoising techniques while impulsive characteristics are also well retained. The envelope spectrum of the signal based on NLM denoising is shown in figure 8 where the rotating frequency f_r , and the defect-induced frequency BPFO at 90.86Hz as well as its multiplication frequencies are clearly identified.

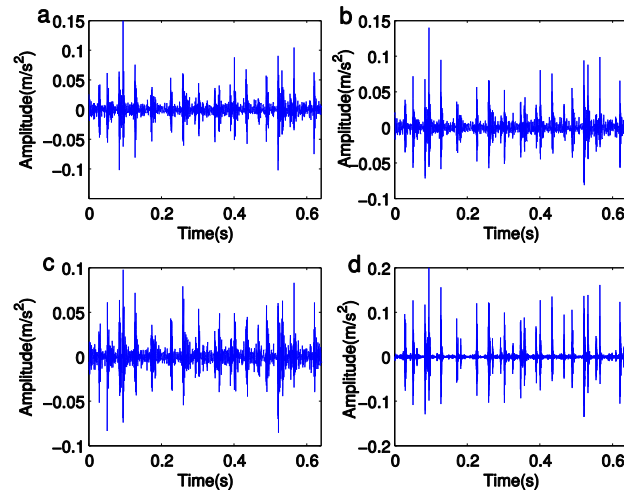


Figure 7. The purified signal using (a) WTD (b) TIWD (c) SGWD and (d) NLM denoising

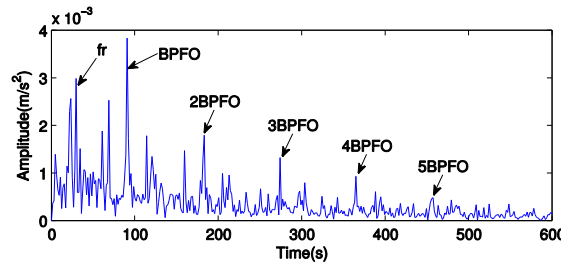


Figure 8. The envelope spectrum analysis results of signal with NLM denoising

4.3. Detecting bearing inner race fault

Figures 9(a) and 9(b) show the original vibration signal and the corresponding frequency spectrums for the bearing with inner race fault. Neither the time domain nor the frequency domain information could tell the fault type.

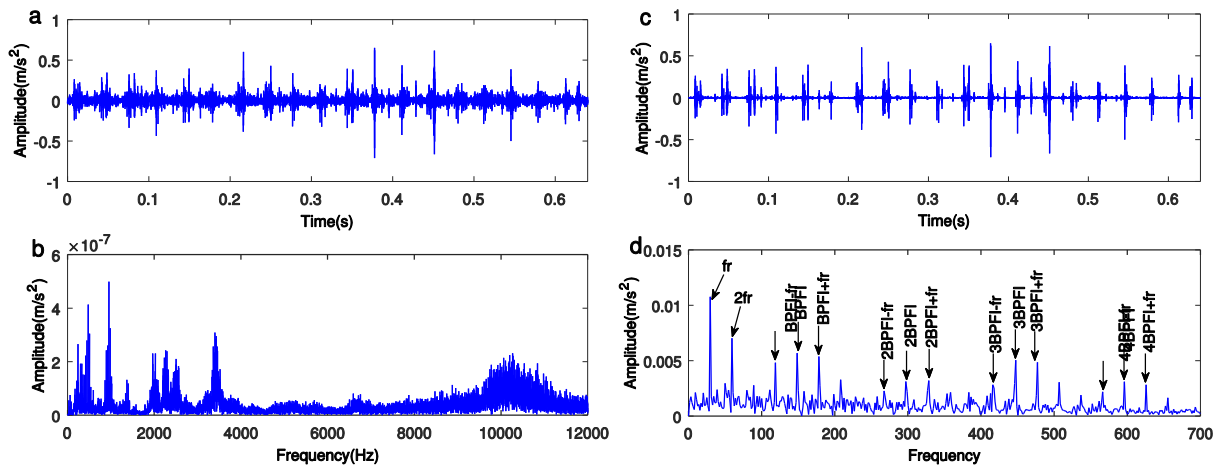


Figure 9. Vibration signal of bearing with an inner race fault time domain waveform, (b) FFT spectrum, (c) the purified signal with NLM denoising, (d) the envelope spectrum of the denoised signal

The purified signal based on NLM method is shown in figure 9(c), and ambient noises are eliminated and weak impulsive signatures are also well preserved. Vibration signal of inner race fault is an amplitude modulated waveform, thus side bands are expected at two frequencies, i.e. $BPFI \pm f_r$. The envelope spectrum of the purified signal is shown in figure 9(d). It can be seen that BPFI, 2BPFI, 3BPFI and 4BPFI along with the side bands ($148.6 \pm 29.87\text{Hz}$, $297.3 \pm 29.87\text{Hz}$, $447.4 \pm 29.87\text{Hz}$, $596.36 \pm 29.87\text{Hz}$) are all prominent. It well reveals that there exists an inner race fault in the bearing.

4.4. Detecting bearing rolling element fault

The vibration signal of bearing rolling element fault is also introduced to testify the validity of proposed method. It presents aperiodic the impulsive features, for the reason that the defect point processed on the rolling element surface sometime may not contact with the raceway. The raw signal and frequency spectrum are displayed in figures 10(a) and 10(b). The purified signal based on NLM method and its corresponding envelope spectrum are shown figures 10(c) and 10(d). From figure 10(d), we can see that the shaft rotating frequency 39.89Hz, the fault frequencies (BSF=79.79Hz, 2BSF=159.6Hz) are detected and then the rolling element fault will be definitely detected. Through the above experimental investigations, it can be found that NLM denoising technique can effectively remove noises and well recover fault features of defective bearings.

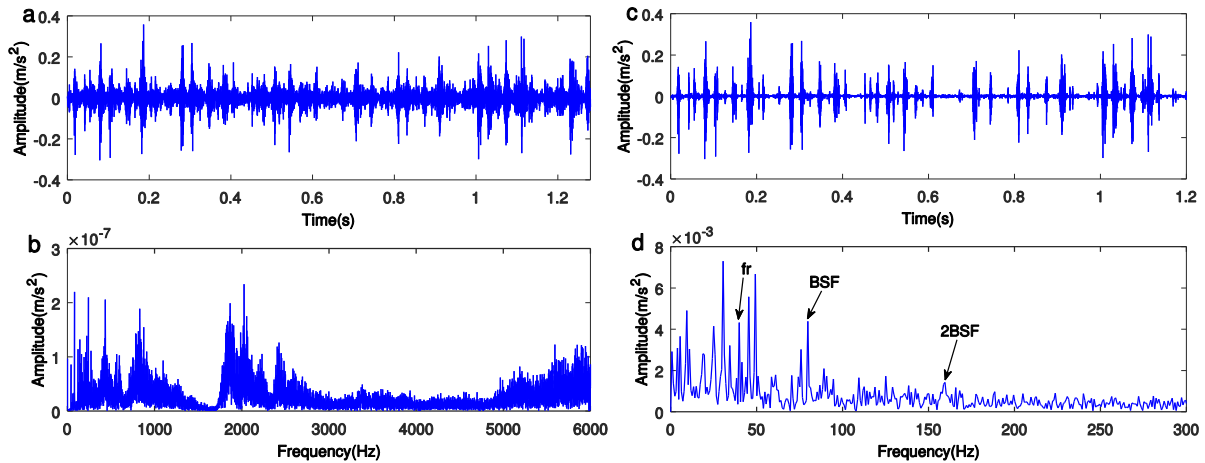


Figure 10. Vibration signal of bearing with an outer race fault (a) time domain waveform, (b) FFT spectrum, (c) the purified signal with NLM denoising, (d) the envelope spectrum of the denoised signal

5. CONCLUSIONS

In this paper, NLM denoising technique is adopted to the bearing fault detection, which on the one hand can remove noise through the weighted average of the similar patches, one the other hand preserved the weak impulsive features. A parameter selecting method of NLM algorithm is proposed in this work, whose effectiveness is demonstrated with simulation analysis in comparison with WTD, TIW, and SGWD. It has been demonstrated that NLM denoising achieves higher SNR values with the lower RMSE and PDR values for all the simulated signals. Moreover, Hilbert envelope spectrum combined with NLM denoising is applied to extract impulsive signatures of bearing vibration signals. The proposed approach is evaluated via bearing vibration signals with three kinds of faults. Results well demonstrate that the present approach is feasible and effective to detect defects in rolling bearing.

ACKNOWLEDGMENT

The financial sponsorship from the project of National Natural Science Foundation of China (51475098 and 61463010) and Guangxi Natural Science Foundation (2016GXNSFFA380008) are gratefully acknowledged. It's also sponsored by Guangxi key laboratory of manufacturing system & advanced manufacturing technology (15-140-30-001Z).

REFERENCES

- [1] R. B. Randall, J. Antoni Rolling element bearing diagnostics—a tutorial [J]. *Mechanical Systems and Signal Processing*, 2011, 25(2): 485-520.
- [2] R. Yan, R. X. Gao Hilbert–Huang transform-based vibration signal analysis for machine health monitoring [J]. *Instrumentation and measurement, IEEE Transactions on*, 2006, 55(6), pp. 2320-2329
- [3] Y. Wang, M. Liang An adaptive SK technique and its application for fault detection of rolling element bearings [J]. *Mechanical Systems and Signal Processing*, 2011, 25(5), pp.1750-1764.
- [4] R. Yan, R. X. Gao, X. Chen Wavelets for fault diagnosis of rotary machines: A review with applications [J]. *Signal Processing*, 2014, 96, pp. 1-15.
- [5] H. Qiu, J. Lee, J. Lin. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics [J]. *Journal of sound and vibration*, 2006, 289(4), pp.1066-1090.
- [6] J. Choi Translation-invariant Wavelet Denoising Method Based on a New Thresholding Function for Underwater Acoustic Measurement [J]. *Transactions of The Korean Society for Noise and Vibration Engineering*, 2006, 16(11), pp. 1149-1157.
- [7] W. Sweldens The lifting scheme: A construction of second generation wavelets [J]. *SIAM Journal on Mathematical Analysis*, 1998, 29(2), pp. 511-546.
- [8] R. Zhou, W. Bao, N. Li. Mechanical equipment fault diagnosis based on redundant second generation wavelet packet transform [J]. *Digital signal processing*, 2010, 20(1), pp.276-288.

- [9] L. Shao, R. Yan, X. Li. From heuristic optimization to dictionary learning: a review and comprehensive comparison of image denoising algorithms [J]. IEEE Transactions on Cybernetics, 2014, 44(7), pp. 1001-1013.
- [10] B. H. Tracey, E. L. Miller. Nonlocal means denoising of ECG signals [J]. IEEE Transactions on Biomedical Engineering, 2012, 59(9), pp. 2383-2386.
- [11] A. Buades, B. Coll, J. M. Morel. A review of image denoising algorithms, with a new one [J]. Multiscale Modeling & Simulation, 2005, 4(2), pp.490-530.
- [12] B. Huhle, T. Schairer, P. Jenke. Fusion of range and color images for denoising and resolution enhancement with a non-local filter [J]. Computer vision and image understanding, 2010, 114(12), pp.1336-1345.
- [13] T. Teuber, A. Lang. A new similarity measure for nonlocal filtering in the presence of multiplicative noise [J]. Computational Statistics & Data Analysis, 2012, 56(12), pp. 3821-3842.
- [14] V. D. Ville D, M. Kocher SURE-based non-local means [J]. IEEE Signal Processing Letters, 2009, 16(11) pp. 973-976.
- [15] V. D. Ville D, M. Kocher. Nonlocal means with dimensionality reduction and SURE-based parameter selection [J]. IEEE Transactions on image processing, 2011, 20(9), pp. 2683-2690.
- [16] M. Feldman Hilbert transform in vibration analysis [J]. Mechanical systems and signal processing, 2011, 25(3), pp. 735-802.
- [17] D. L. Donoho, J. M. Johnstone Ideal spatial adaptation by wavelet shrinkage [J]. Biometrika, 1994, 81(3), pp. 425-455.
MFS-MG User Guides, SpectraQuest Inc, Richmond, 2012

The Influence of B -value Distribution in Damage Evaluation of Structural Material Based on AE Parametric Analysis

M. T. I. Khan, S. Tsurumaru and R. Hidaka

Graduate School of Science and Engineering, Saga University, 1 Honjo, Saga 840-8502, Japan

ABSTRACT

Large number of acoustic emission (AE) events with varying amplitudes are generated in brittle failure of structural materials. A consequent method of describing the distribution characteristics of amplitudes is emphasized by describing their b -values, which is widely defined as slope of the log-linear frequency magnitude distribution of AE events. Conventionally, the b -value of AE is computed by using the methods adopted in seismology since many similarities exist between AE and seismic waves. The b -value is usually obtained either from the cumulative frequency distribution or the discrete frequency distribution of earthquake magnitudes and the same principles hold good for the AE data analysis also. Therefore, b -value characterized by the Gutenberg and Richter's relations have been investigated according to the Aki's modification for AE events generated in material cracks or crack propagations towards final damage of structural materials in the present research. Aki's proposed maximum likelihood method for calculating b -values have been applied from AE maximum amplitude distribution for acquired AE events during the crack damage experiments in tensile loading to ductile iron (pearlite). The calculation shows good evidence about crack initiation, propagation and damage phenomena as well comparing with microstructural video data.

Keywords: AE monitoring, b -value calculation, Aki's Maximum likelihood method, Structural material, Damage evaluation.

Corresponding author: Khan M. T. I. (email: khan@me.saga-u.ac.jp)

1. INTRODUCTION

Brittle failure occurs inside structural material under the formation of micro cracks and the growth of micro cracks towards the matured cracks due to the tensile loads, compression loads, shear loads etc. until the complete failure. Elastic waves generate inside the material with the occurrence of minute defects like cracks or micro-cracks and propagate through the materials. Acoustic emission (AE) is a nondestructive measurement technique which acquires the elastic energy to generate acoustic waves and clarifies the damage condition inside the material through several defined parameters, known as AE parameters. AE sensor responds to dynamic motion due to damage inside the material and converts the mechanical movement into an electric voltage signals, which is visualized as an AE event. Each AE event can cause the acquisition system to record multiple hits with numerous channels, in which, the received hit generates AE signals exceeding the pre-defined threshold voltage. The maximum voltage in a waveform of one hit indicates the maximum amplitude, generally, is expressed in decibels (dB). As an important parameter, it determines detectability of the signal. However, amplitude below the operator defined threshold voltage does not count in signal-amplitude analysis. As this process stops for a predetermined amount of time, the hit is finished. AE data from the hit is then received to the computer for the desired parametric analysis [1~3].

Materials undergoing fracture generates AE waves with varying amplitude distribution with linear feature and descending gradient. The slope of the logarithmic linear distribution of AE event-magnitude is defined as b -value. The b -value has general feature to show the natural projection to unobserved future trend [4]. The general objective of b -value calculation is to explore the theoretical frequency-magnitude coefficient, b -value with number of fractal datasets in earthquake or similar bursting signals during material fractures. In seismology, widely used the Gutenberg-Richter relationship for analyzing the cumulative frequency

distribution data [5] and the Aki's method for the analysis of discrete frequency distribution data [6] are still considered as the best fitting in magnitude distribution prediction based on the basis of b-value calculation procedures [7].

Accordingly, the damage propagation data of AE events (hits) in structural materials fracture are considered to be analyzed with the similar high sensitivity of damage prediction like in seismology by b-value calculation. Therefore, in the present research the fracture process (crack propagation behavior) of a ductile iron plate (pearlite) has been experimentally evaluated by calculating the b-value distribution of amplitude parameters in cumulative as well as discrete AE events distribution under the tensile loadings.

2. EXPERIMENTAL METHODOLOGY

2.1. Specimen preparation for the experiment

Specimen for the desired experiments ductile cast iron (pearlite) was prepared from a cast iron block with the dimension of $100 \times 80 \times 260\text{mm}$. Chemical compositions of the test materials are listed in table 1. The numerical values in this table are represented by the corresponding values in percent content of carbon (C), silicon (Si), manganese (Mn), phosphorus (P), sulphur (S) and copper (Cu) in percent mass respectively. Mechanical properties of pearlite ductile cast iron is shown in table 2. The white cast iron is heat treated according to the method of pearlite grain determination test as Japanese Industrial Standards (JIS G 0552) for converting the hard iron carbide structure into malleable pearlite malleable matrix. The pearlite structure contains some combined carbons in the matrix which result higher strength and hardness. The specimen has been prepared by cutting the test material from the bottom part of the cast block. Then the central part of the specimen has been mirror polished to examine the crack propagation by conducting the microstructural analysis as well. The dimensions of the prepared specimen with AE sensors are shown in figure 1.

Table 1: Chemical composition of material (PDI) used in AE experiment.

| | C | Si | Mn | P | S | Cu |
|----------|------|------|------|-------|-------|-------|
| Mass (%) | 3.68 | 2.08 | 0.16 | 0.017 | 0.001 | 0.665 |

Table 2: Mechanical properties of material (PDI) used in AE experiment.

| Tensile strength (Mpa) | 0.2% Proof stress (Mpa) | Stretching (%) | Hardness (HBW) |
|------------------------|-------------------------|----------------|----------------|
| 793 | 483 | 5.9 | 218 |

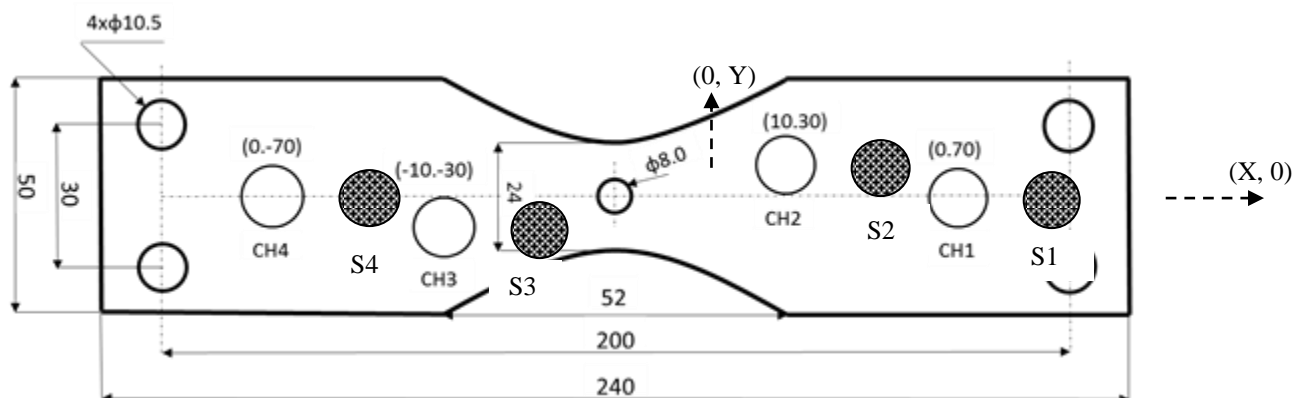


Figure 1. Dimension of specimen with 4 AE sensors (S1-S4).

2.2. Experimental procedure

According to the objective, the test specimen was attached to a tensile testing machine (Autograph) for conducting tensile test with required tensile loading which created the adequate cracks to the specimen until occurring the final crack-damage. Four AE sensors were attached to the specimen by inserting the special coupling silica gel between the sensor and specimen for avoiding the mismatch of the acoustic impedance. Each sensor is connected to the pre-amplifiers and then to the 4-channel main AE amplifier.

Under the tensile loading, structural deformation was generated inside the material and therefore, AE data (AE event) were received through four AE sensors (R15 α ; physical acoustics Ltd.) together. After applying adequate filtration, the AE signals are stored in a digital storage oscilloscope (DSO) from where the data were transferred to the personal computer for the analysis. For microstructural analysis, high resolution (200 times zoom) video images of crack initiation and propagation until crack damage were taken continuously by a digital video microscope. The panoramic view of the experimental set-up is shown in figure 2. The schematic view of the experimental procedure is shown in figure 3 as follows.

AE data for total stages of crack propagation were saved in personal computer, from where b-value distribution was evaluated and clarified according to the Gutenberg-Richter relationship and the Aki's method.

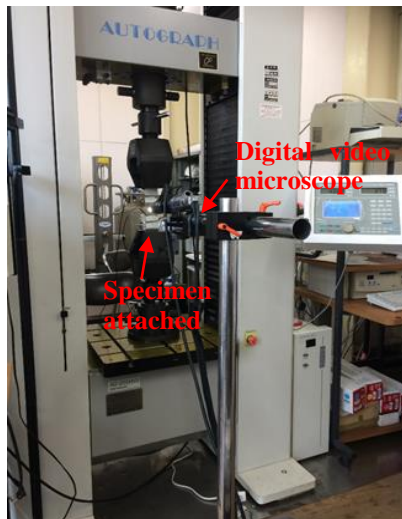


Figure 2. A panoramic view of the set up.

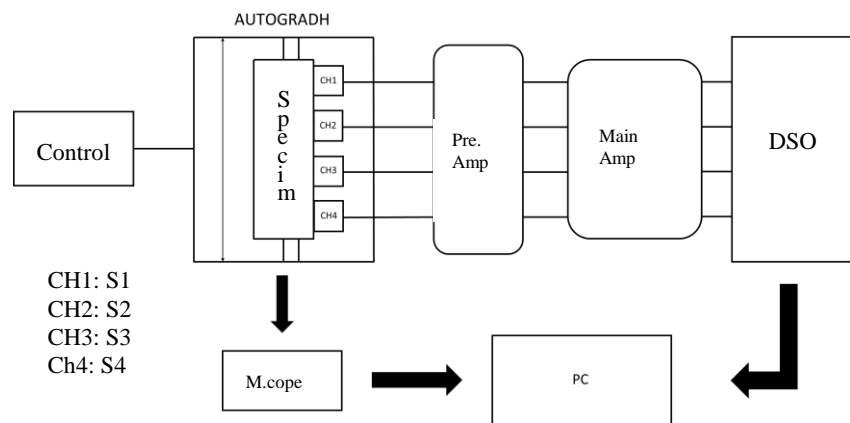


Figure 3. Schematics of the experimental procedure.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. b-value calculation by Gutenberg-Richter Relationship and Aki's method

The calculation of b-value based on the Gutenberg-Richter Relationship was performed by applying the following equation.

$$\log_{10} N = a - b(A_{dB} / 20) \quad (1)$$

Where, N was indicated as AE counts (number of AE amplitude crossed the threshold value) with amplitude greater than the defined threshold, A_{dB} was indicated as the maximum amplitude in dB, a was indicated as the empirical constant in b-value constant with 20 correction factor.

Similarly, the b-value based on the Aki's maximum likelihood discrete frequency method was calculated also to compare the crack propagation technique with Gutenberg-Richter relations. The Aki's method is explained as follows.

$$b = [20 \log_{10} e] \div [<a> - a_c] \quad (2)$$

Where, b was indicated as the AE b -value, e was indicated as an empirical constant (2.71828), $<m>$ was indicated as the average magnitude and m_0 was indicated as the defined threshold value in AE data acquisition. In both calculation as mentioned above, impact is given in AE maximum amplitude distribution function. However, as the Gutenberg-Richter relation emphasized on cumulative distribution function, the b -value magnitude was focused to a magnified scale compared to the Aki's discrete likelihood method. Both of results are compared as follows according to the total acquisition AE data (hits). The comparison result is shown in figure 4.

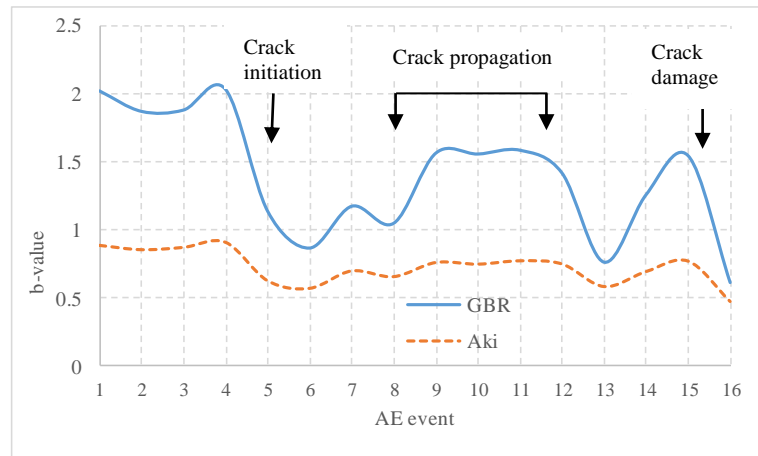


Figure 4. Mean line of b -value by Aki's method and Gutenberg-Richter(GBR) method.

The crack propagation analysis by b -value calculation shown in figure 4 clearly shows the crack initiation stage, crack propagation stage as well as crack damage stage. In both calculation methods of b -value, it is found that the b -value is inversely proportional to the maximum amplitude distribution of AE event. Therefore, when the crack was initiated in the material at hit number 5, AE amplitude was increased to a maximum level, and thus, the b -value was decreased. According to the material fractal theory, the starting stage of material deformation or cracking needs higher energy compared to its propagation stage. Therefore, AE energy (function of AE amplitude) release in propagation stage is lesser than that of the initial stage. Therefore, in the experimental results, the crack propagation stage showed the lesser b -value distribution compared to its starting calculations (hits 1~4), however, it indicated the higher value compared with the crack initiation stage (hit 5). Similar result was also found in crack damage stage as well. These crack propagation behaviour was clarified by both calculation methods of b -value.

Again the above mentioned cracking stages of the pearlite specimen under tensile loading were verified with the microstructural images. A high vision video microscopic (KEYENCE, Japan) camera was used and the crack propagation was recorded under 200 times zooming images. Thus, the images of crack initiation and propagation until crack damage were documented as follows (figure 5).

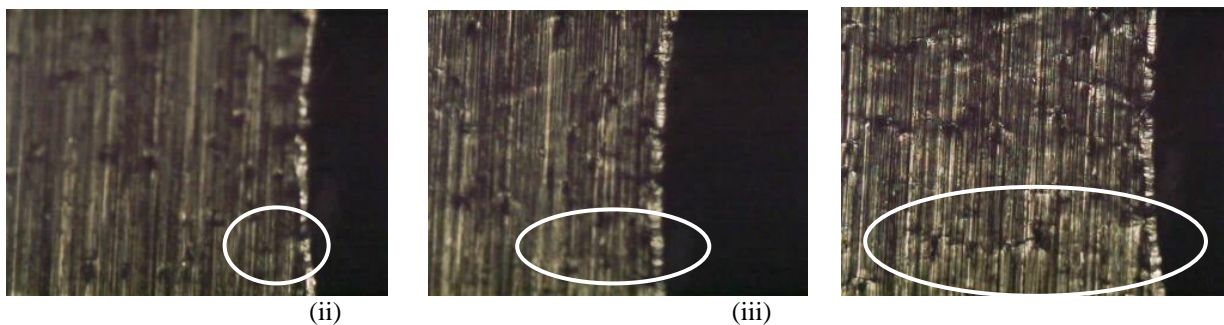


Figure 5. Microscopic images of (i) crack initiation (hit 5), (ii) crack propagation (hit 12), (iii) matured crack (hit 13).

In microscopic images, crack initiation as well as propagation stages were clearly identified. Although, several other cracks were also found, however, the mentioned crack was identified as the crack of final damage. Furthermore, due to multiple matured cracks, final damage stage was obscured in the present experiment. It is thought, as the cracking area of the specimen was considered very small (20 mm), before growing the long unique matured crack, complete damage of the material occurred.

4. CONCLUSIONS

The crack propagation behavior in a plate-type ductile cast iron (pearlite) specimen was successfully identified in the present research by the b-value analysis. Both Gutenberg-Richter relationship and Aki's maximum likelihood method in b-value analysis indicated the crack initiation stage, crack progression stage as well as damaging stage clearly. The selection of cumulative AE amplitude distribution function in Gutenberg-Richter technique indicates the variation in b-value distribution clearly compared with the Aki's method of discrete amplitude selection method. However, both results were clearly satisfied in the present experiment.

Although, mentioned b-value monitoring methods are widely used in earthquake analysis, however, present results showed that material damages (tensile loading in present experiment) can also be clarified by b-value analysis in AE monitoring methods. For getting clear microstructural images of crack growth behavior, although, much wider cracking area or path in the specimen was needed in deep understanding of crack growth and crack damage behaviors along with AE technique in b-value analysis.

REFERENCES

- [1] Grosse CU, Ohtsu M. Acoustic Emission Testing. Heidelberg: Springer; 2008.
- [2] Ohtsu M. Recommendation of RILEM TC 212-ACD: Acoustic emission and related NDE techniques for crack detection and damage evaluation in concrete: test method for classification of active cracks in concrete structures by acoustic emission. 2010 Mater Struct., 43 (9), 1177-81.
- [3] Carpinteri A., Lacidogna G., Accornero F., Mpalaskas A.C, Matikas T.E. Aggelis D.G., Influence of damage in the acoustic emission parameters (2013) Cement & Concrete Composites, 44, 9-16.
- [4] Gutenberg B. & Richter C F. (2nd Ed) (1954) Seismicity of the Earth and Associated Phenomena, Princeton University Press, Princeton, 310.
- [5] Gordon Lan and Wittes Janet, The b-value: A tool for monitoring data, (1988) Biometrics 44, 579-585.
- [6] Rao MVMS & Prasanna K JL. Amplitude distribution analysis of acoustic emissions and investigation of the development of brittle fracture in rock (2006) Indian Journal of Pure & Applied Physics 44, 820-825.
- [7] Emile A. Okal, Barbara A. R. On the variation of b-values with earthquake size, (1994) Physics of the Earth and Planetary Interiors, 87, 55-76.

A Self-Adaptive AE Signal Classification Method on Energy Domain

Yong Zhou¹, Da Wang¹, Li Lin²

¹ School of Software, Dalian University of Technology, Dalian, China

² Department of Traffic and Transportation Engineering, Dalian Jiaotong University, Dalian, China

ABSTRACT

This paper presents a self-adaptive data analysis method on acoustic emission (AE) signals which are generated by the fracture test on the rotation shaft. When collecting AE signals from the fracture test, it is easy to get some kinds of noise signals, such as lead-off signal and whacked signal. And considering the high sampling rate on AE signal collection, it is necessary to find an effective method to intelligently pre-process this huge data. So this paper uses the empirical mode decomposition (EMD) to find the particular features of the AE signal. The proposed analysis method consists of feature extraction using the EMD method and fault classification using the artificial neural network (ANN) technique. Using EMD method to decompose the sample signals, it obtains some intrinsic mode function (IMF) components. IMFs can be used to calculate a carefully designed feature vector which contains energy entropy, energy distribution and total energy. In the fault classification, these energy feature vectors of different kinds of signals are used as inputs to train the ANN. Experimental results indicated that this analysis method can accurately identify the AE signals caused by the fracture test on rotation shafts from some kinds of noise signals produced at the same time. So this analysis method can be used in the acoustic emissions collection and identification.

Keywords: Empirical mode decomposition; Elman neural network; Feature extraction; Fault diagnosis; Energy entropy.

Corresponding author: Yong Zhou (email: kevinzh@dlut.edu.cn)

1. INTRODUCTION

The AE signal analysis plays an important role in the field of digital signal processing and fault diagnosis. In general, the mechanical system always works with the acoustic emissions when the system goes to failure. In the process of acoustic emission nondestructive testing, the acoustic emission is sensitive to the material and is easy to be interfered by mechanical and electrical noise. Therefore, rich experience are necessary to interpret the collected AE data-sets. As a result, it is significant to identify the collected AE signals and find the noise signals from the collected signals. Because of the high frequency of the AE signals, the sampling rate could reach about 1000kHz. And during the fracture test, the AE phenomena only happens a few times. So it is important to get effective information, the AE signal, from the huge ineffective signals and interference signals. The identification of AE signals is a pattern recognition problem essentially. It needs effective feature extraction and applies efficient and accurate classifier as well to improve the accuracy of AE signal processing and fault diagnosis.

Feature extraction is one of the most critical obstacle in the acoustic emission diagnostic procedure. By using some transformation functions, it is a commonly used method to transform the data in the time domain into the frequency domain in signal analysis. And the Fast Fourier transform (FFT) is one of the most useful and convenient method for frequency domain feature extraction in mechanical system diagnosis. For example, using FFT algorithm to pre-estimate the frequencies of a signal make it easy to detect the rotor broken-bar faults feature component hidden behind the strong supply frequency component in the spectrum of the stator current [1]. The Short Time Fourier Transform (STFT) belongs to an improved Fourier-related transform. According to the uncertainty principle by W.Heisenberg, the resolution in the time domain and in the frequency domain are mutual restricted. They cannot achieve the best at once.

The wavelet transform (WT) can provide an adjustable window function and give a good resolution in both the time and frequency domains. The higher the frequency is, the shorter the window size become. Therefore, the WT has been applied in various applications for signal feature extraction and fault diagnosis. For example, the continuous wavelet transform (CWT) was selected for extracting efficient features of the surface electromyography signals in the research [2]. CWT includes time-based information as well as frequency-based, making muscle diagnosis easier. An algorithm based on a combination of the discrete wavelet transforms (DWT) and neural networks is used to detect and classify the internal faults in a two-winding three-phase transformer[3]. The wavelet packet decomposition(WPD) extracts the instantaneous power as features to diagnose the motor fault. Several wavelet transform analysis methods mentioned above are widely used in various fields of signal analysis with good results[4].

Although the WT shows a good effect on the signal processing, the algorithm should choose a suitable wavelet function before analysis. And different wavelet functions have different results on WT. As a result, a kind of new method called empirical mode decomposition (EMD) was first introduced in 1998 [5]. It is a self-adaptive analysis method and good at analyzing the non-linear and non-stationary signals. This method mainly has two steps. The EMD method decomposes signals into some intrinsic mode functions(IMF) based on the local time domain signals themselves. Then, the IMFs are transformed by Hilbert-Huang Transform (HHT) to get the frequency and energy domain information of the signal. As a result, the EMD method is used to analysis the group of vibration signals of varied faults on the experimental gearboxes with high accuracy and speed [6]. Except for the frequency domain information, the energy domain information also plays an important role on fault diagnosis. The EMD energy entropy mean is computed to diagnose the faults on the ball bearing [7].

The AE signal is a kind of non-linear and non-stationary signal. Traditional signal analysis methods are not suitable for this kind of signal. AE signals generated from plywood damage are decomposed by EMD and transformed by HHT to extracted features of four types of plywood damage signals [8]. AE signals caused by partial discharge(PD) are transformed to a 3D Hilbert spectrum using the EMD method. And several defect types of PD are diagnosed correctly[9]. So the EMD method is an effective way to analyze the AE signal, but there are only a few researches and applications on it.

In recent years, the neural network (NN) has become the most popular method in pattern recognition. So the EMD method and NN always work together to forecast the data and diagnose the faults. Several damage features always overlap in one signal. The EMD method can decompose this kind of signal into some IMFs and one IMF contains part damage features of all. Then a 3D Hilbert spectrum can be computed by HHT and some useful features can be extracted from it. NN learns and recognizes different kinds of signals finally. A modified EMD-FNN model is proposed for forecasting the mean monthly and daily wind speed of Zhangye of China [10]. And an EMD-RBF model is used to forecast precipitation with good results[11]. In the fault diagnosis field, to the instability of AE signal of rock fracture, the method combining AE parameters, EMD and BP neural network is introduced[12]. And a flow regime identification method using the EMD combined with Elman neural network is used on the gas-liquid two-phase flow [13].

In this study, an EMD-ENN model that hybridizes EMD and ENN is proposed to identify AE signals [14]. First, the AE signal data-sets are decomposed into a collection of IMFs and a residue by EMD. Second, both the IMF components and the residue are transformed by HHT to get the energy domain features. Then, these features are used to train the ENN model. Finally, this trained ENN model are used to identify different kinds of signals. And the result shows that this model can identify the AE signals from different kinds of noise signals effectively.

2. PROPOSED APPROACH

2.1. Algorithm Process

In this paper, an EMD-ENN model, which combines EMD with ENN, is proposed for the AE signal identification. The methods used in this approach are briefly introduced below. And the flowchart which is used to describe the algorithm is shown in figure 1.

- Step 1: Use the EMD to decompose the original signal data-sets into six IMFs and a residue, which have simple frequency components and are relatively easy to analysis.
- Step 2: Calculate the EMD energy distribution in the whole of the corresponding IMFs. Then calculate the energy entropy and interval average energy of the original signal data-sets.
- Step 3: Construct the ENN models for the original signal data-sets, and apply the AE signals and noise signals to train the established ENN models.
- Step 4: Use other samples of the original signal data-sets to detect the trained ENN models, and evaluate the EMD-ENN model.

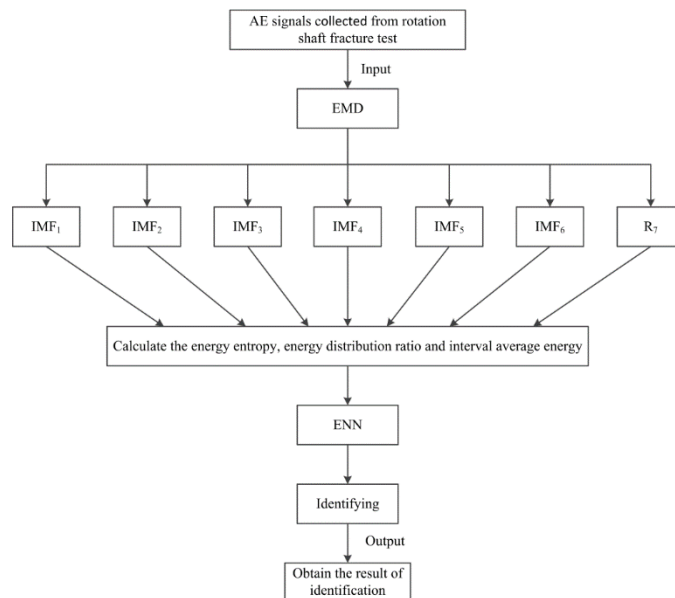


Figure 1. EMD-ENN flowchart

2.2. Features Extraction with EMD Method

EMD is developed from the simple assumption that any signal is consisted by different number of IMFs, which means different kinds of simple mode of oscillations. This adaptive decomposition method is usually used to analysis non-linear and non-stationary signals such as the AE signals.

Each IMF mainly has two features:

- 1) The number of extrema and the number of zero crossings must either equal or differ at most by one in the entire signal.
- 2) The mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at any point.

With the use of the EMD method, an AE signal can be decomposed into six IMFs and a residue.

After decomposing the AE signals into six IMFs and a residue, the instantaneous amplitudes of each IMF can be computed by HHT as (1).

$$a_i(t) = \text{hilbert}(f_i(t)) \quad (1)$$

where $f_i(t)$ is the i th IMF, and $a_i(t)$ is the instantaneous amplitudes of IMF _{i} .

The interval average energy of each IMF is calculated as (2):

$$E_i = \sum_{j=1}^{4096} |a_{ij}|^2 \quad (2)$$

where i means the instantaneous amplitudes of IMF _{i} and j means the j th sampling point in $a(t)$. And then, calculate the interval average energy of all IMFs as (3).

$$E = \sum_{i=1}^7 E_i \quad (3)$$

And then, the EMD energy entropy is calculated as (4):

$$H_{en} = -\sum_{i=1}^7 p_i \log(p_i) \quad (4)$$

where $p_i = E_i/E$ is the percentage of energy distributed in IMF _{i} to the total energy of the signal[15].

| |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Input: a sample $s(t)$ |
| Output: energy domain features $F[9]$ |
| <pre> 1: Procedure feature_extraction() 2: $i \leftarrow 1$, and $j \leftarrow 1$ 3: call limited_EMD($s(t)$) to get the IMF $f[j]$, and $total_energy \leftarrow 0$ 4: for $i \leq 7$ 5: $hil[i] \leftarrow \text{hilbert}(f[i])$ 6: Compute the instantaneous amplitudes $amp[i]$ by $hil[i]$ 7: Compute the energy $ener[i]$ of $f[i]$ by $amp[i]$ 8: $total_energy \leftarrow total_energy + ener[i]$ 9: end 10: for $j \leq 7$ 11: $F[j+1] \leftarrow ener[j]/total_energy$ 12: end 13: $F[9] \leftarrow total_energy$ 14: Compute the energy entropy ee of $s(t)$ 15: $F[1] \leftarrow ee$ 16: end procedure </pre> |

Table 1: Energy domain feature extraction algorithm.

Finally, by following the pseudo code in table 1, the energy feature vector can be computed which contains 9 parameters. Parameter 1 is the energy entropy, which shows the even degree of energy distribution. While parameter 2 to 8 shows the energy distribution on each IMF. And parameter 9 shows the total energy that the signal has. By using this three kinds of parameters, almost all the energy information calculated by the EMD is described in detail by the vector presented in this paper.

2.3. Elman Neural Network Structure

The ENN is a simple recurrent neural network which was introduced by Elman in 1990. This type of network is composed of nodes arranged in input, hidden and output layers.

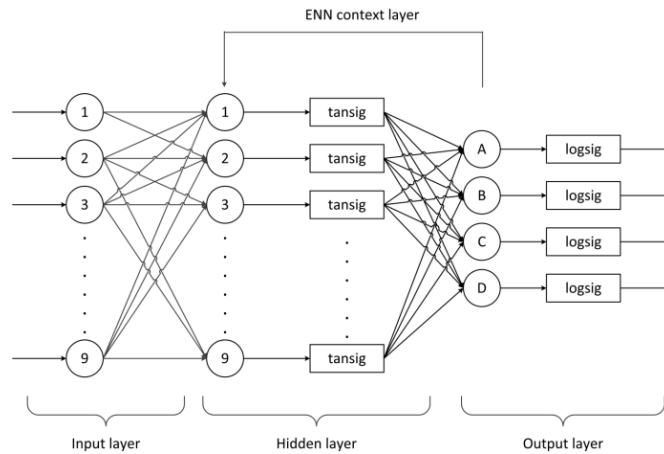


Figure 2. ENN structure used in this paper

To some degree, it is similar to a three-layer feed-forward neural network. However, it has a context layer that feeds back the hidden layer outputs in the last steps. This context layer allows ENN to form a sort of short-term memory. In figure 2, node1 inputs the feature energy entropy. Node2 to node8 input the feature energy distribution ratio of IMF1 to IMF7. And node9 input the feature interval average energy. And the ENN with a single hidden layer is selected in this paper. Since there are four kinds of signals to recognize, four neurons are used in the output layer, where each output of four neurons represents the matching degree of the corresponding signal. Node A represents the lead-break signal. Node B represents the whacked signal. Node C represents the noise signal. Node D represents the AE signal.

3. EXPERIMENTAL RESULTS

3.1. Experimental Setup

The experimental apparatuses include a rotation shaft fracture test platform, an AE sensor, a pre-amplifier, a signal separator and an AE signals collection system. The system collects the AE signals until the rotation shaft goes to failure when doing the fracture test on it. And the sampling frequency is set as 1000000. This system collects the AE signals during the fracture test. But at the same time, some noise signals can also be collected such as the whacked signal. And this kind of noise signal influences the quality of the collected AE signal. In this paper, three kinds of noise signals were made to collect.

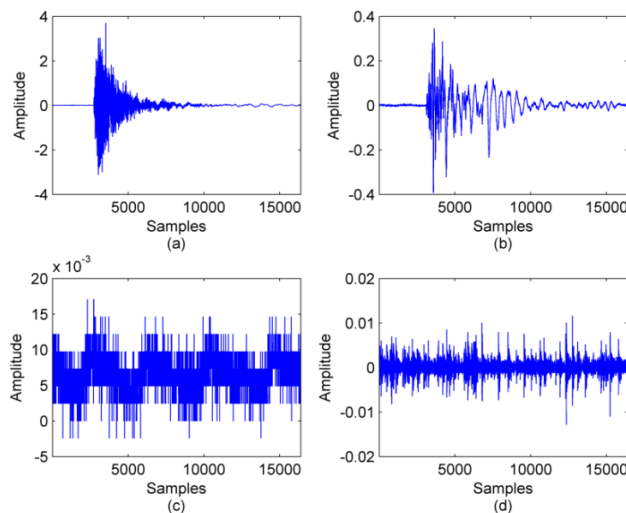


Figure 3. Wave-form of four signals. (a) lead-break signal; (b) whacked signal; (c) noise signal; (d) AE signal

To collect the normal AE signal, the test is giving pressure to the working rotation shaft until it runs to failure. To collect the lead-break signal, the test is doing the lead breaking experiment when the rotation shaft works. And to collect the whacked signal, the test is using something hard like a pen or hammer to knock the rotation shaft when it works. At last, to collect the environmental noise signal, the test is collecting the signals when the rotation shaft stops to work and the experiment remains unchanged.

Figure 3 displays the wave-forms of four kinds of collected signals. The x-axis represents each sampling points of the signal and the y-axis represents the amplitude of each sampling point.

3.2. The Experimental Process

The features of four kinds of signals are explored. 20 samples are collected from each signal. The analysis of four signals is based on these 80 samples. As shown in figure 4,

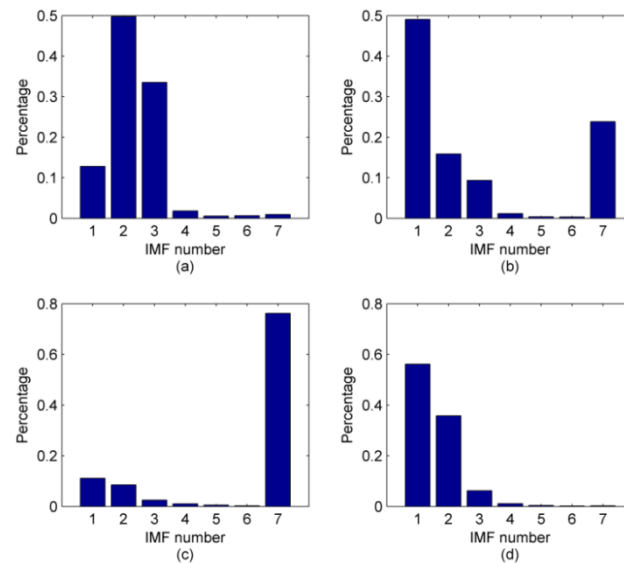


Figure 4. Energy distribution ratio of four signals. (a) lead-break signal; (b) whacked signal; (c) noise signal; (d) AE signal

It displays the wave-forms of four signals. The x-axis represents each sampling points of the signal and the y-axis represents the amplitude of each sampling point. Then, to further analyze the sample signals, the limited EMD method which decomposes these signals into seven components.

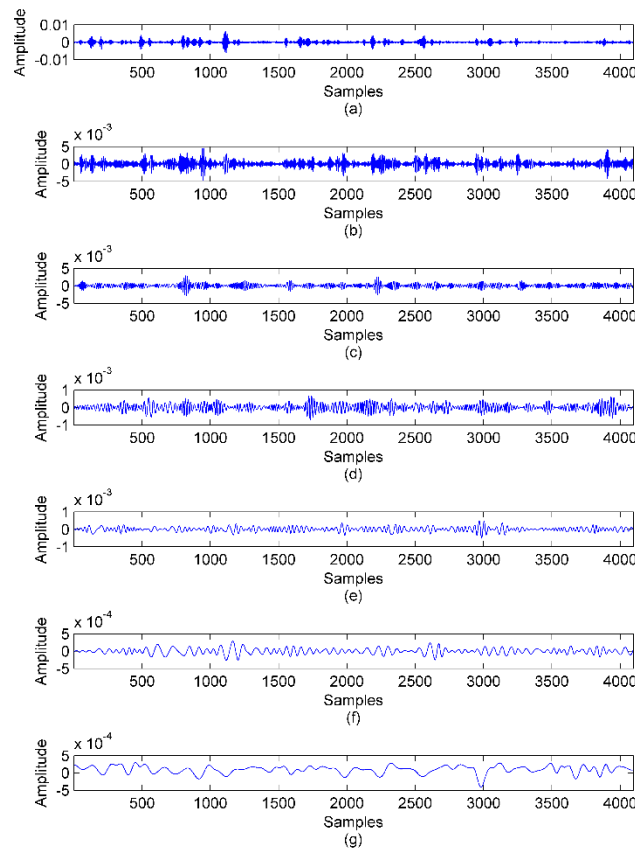


Figure 5. Results of decomposition of AE signal. (a) IMF1; (b) IMF2; (c) IMF3; (d) IMF4; (e) IMF5; (f) IMF6; (g) residue.

Figure 5 shows the results of decomposition on the AE signal. In the figure, IMF1 to IMF6 represent the varied frequencies from high to low. The IMF7 contains the rest of the frequencies and the residue.

According to the formula given in the previous section 2.2, the energy feature of four kinds of signals is extracted. Figure 4 shows the energy distribution of four kinds of signals. The x-axis represents the IMF components and the y-axis represents the EMD energy entropy in the whole of the corresponding IMF. It can be seen that the distribution of signal energy in four kinds of signals have different regularity.

Table 1. Energy entropy of every sample and mean value of each kind of signal

| Sample number | Acoustic emission signal | Lead- break signal | Knock signal | Noise signal |
|---------------|--------------------------|--------------------|--------------|--------------|
| 2 | 1.4631 | 1.6649 | 1.8594 | 1.3942 |
| 3 | 1.5504 | 1.9907 | 1.5412 | 1.3797 |
| 4 | 1.2712 | 2.5382 | 1.6028 | 1.3312 |
| 5 | 1.4271 | 1.9850 | 2.0322 | 1.2479 |
| 6 | 1.4146 | 1.6542 | 2.2512 | 1.2380 |
| 7 | 1.4965 | 2.3650 | 2.5148 | 1.2129 |
| 8 | 1.3935 | 2.4190 | 1.7422 | 2.0418 |
| 9 | 1.4281 | 2.3319 | 2.5067 | 2.0021 |
| 10 | 1.4839 | 2.6093 | 2.2579 | 1.2272 |
| 11 | 1.2969 | 2.0589 | 1.9936 | 2.0456 |
| 12 | 1.3342 | 2.4206 | 2.1413 | 1.8502 |
| 13 | 1.4297 | 2.1159 | 1.9153 | 1.8542 |
| 14 | 1.3752 | 2.6618 | 2.3101 | 1.8153 |
| 15 | 1.3760 | 2.4572 | 1.6487 | 1.7776 |
| 16 | 1.4408 | 2.1046 | 2.3037 | 1.6800 |
| 17 | 1.0987 | 2.5780 | 2.4640 | 1.7076 |
| 18 | 1.3401 | 2.0467 | 1.5130 | 1.2151 |
| 19 | 1.4860 | 2.2132 | 1.8991 | 1.5521 |
| 20 | 1.4090 | 2.2230 | 1.6177 | 1.5512 |
| AVG | 1.3952 | 2.1975 | 1.9993 | 1.5669 |

Table 1 shows the energy entropy of every sample and the mean value of each kind of signal. It can be seen that the energy entropy of noise signals and the AE signals obviously lower than the other two signals. The reason is the energy of noise signal and the emission signal distributes too much in only one IMF.

Table 2: Total energy of every sample and mean value of each kind of signal

| Sample number | acoustic emission signal | lead- break signal | Knock signal | Noise signal |
|---------------|--------------------------|--------------------|--------------|--------------|
| 1 | 0.0249 | 112 | 0.9655 | 0.1979 |
| 2 | 0.0085 | 4445 | 1.2240 | 0.1785 |
| 3 | 0.0082 | 128 | 0.9868 | 0.1781 |
| 4 | 0.0066 | 3440 | 0.8857 | 0.1810 |
| 5 | 0.0040 | 833 | 0.6505 | 0.2002 |
| 6 | 0.0041 | 118 | 0.7944 | 0.2041 |
| 7 | 0.0269 | 613 | 1.4280 | 0.2005 |
| 8 | 0.0087 | 5178 | 0.4361 | 0.0880 |
| 9 | 0.0047 | 2903 | 2.2406 | 0.0921 |
| 10 | 0.0068 | 1589 | 0.6093 | 0.2046 |
| 11 | 0.0105 | 898 | 0.5369 | 0.1091 |
| 12 | 0.0121 | 10367 | 0.8391 | 0.1082 |
| 13 | 0.0234 | 4724 | 0.5201 | 0.1170 |
| 14 | 0.0216 | 38 | 10.700 | 0.1272 |
| 15 | 0.0137 | 436 | 0.3313 | 0.1298 |
| 16 | 0.0225 | 3089 | 1.8525 | 0.1493 |
| 17 | 0.0188 | 4526 | 2.0431 | 0.1533 |
| 18 | 0.0197 | 588 | 0.6501 | 0.2070 |
| 19 | 0.0102 | 128 | 0.4964 | 0.1660 |
| 20 | 0.0188 | 1523 | 0.4080 | 0.1655 |
| AVG | 0.0137 | 2283.8 | 0.9484 | 0.1579 |

Table 2 shows the total energy of every sample and the mean value of each kind of signal. From the perspective of energy distribution, it is obvious that the distribution of acoustic emission signal, lead-break signal and knock signal looks not that different. The energy of these three signals distribute mostly in IMF1 and IMF2. But actually, the total energy released by these three signals differs a lot. With the use of the feature total energy, the success rate of classification on four kinds of signals is increasing obviously.

3.3. Test on the Collected AE signals

In this section, the EMD-ENN method introduced will be used to test a collection of AE signals. Figure 6(a) shows the wave-forms of the AE signal. The x-axis represents the sampling points and the y-axis represents the amplitude of each sampling point. This AE signal is collected from a fracture test on the rotation shaft. The sensor collected all the acoustic emissions until the rotation shaft went to failure during the whole fracture test. And it is obvious that there are two outliers on this collected AE signal. The amplitudes of these two outliers are obviously higher than the AE signal. And the EMD-ENN method will be used to pick these two outliers out.

To pick out the outliers in this AE signal, it is necessary to use a trained ENN model. Then, the collected AE signal is cut into pieces and the length of each piece is set to 16384 sampling points. And compute the feature vector of every piece following the steps introduced already. At last, the trained ENN model is used to identify these feature vectors and the result of the identification can be seen from the node 4 of the output layer.

Figure 6(b) shows the result of the test on the AE signal. The x-axis represents the serious number of the segments of AE signal and the y-axis represents the the matching rate of every piece of the AE signal identified by the ENN. Compared these two pictures, it is obvious that two X marks in picture (b) prove that there are two outliers in corresponding position in figure 6(a). It is a kind of noise signal collected during the collection of the AE signal when doing the fracture test on the rotation shaft. And the test proves that the EMD-ENN model have good effect on the identification of AE signals.

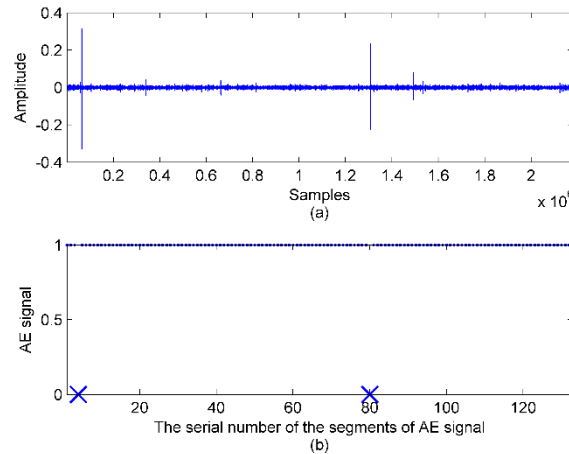


Figure 6. The result of the test on the collected AE signal

4. CONCLUSION

It implemented an AE signal analysis method with some energy domain features calculated to distinguish AE signals from some kinds of ineffective signals when collecting. This method shows an effective way to analyze the non-linear and non-stationary AE signal. The EMD decomposes the collected signals into some IMFs. The sifted IMF components represented useful information of the signal. The energy entropy and distribution correlate closely with each IMF component. The proposed procedure of feature extraction reduced the great dimensions of feature more effectively. With the using of a carefully designed energy feature vector, the ENN could quickly completes the training procedure and obtains a required MSE. The experimental results show the proposed technique performs well for identifying the collected signal types quickly and accurately.

5. ACKNOWLEDGMENT

This work was funded by the National Science Foundation of China under Grant 51275066.

REFERENCES

- [1] Wang, H. X., & Yang, W. D. "Rotor bar fault feature extraction of induction motor base on FFT and MUSIC". Proceedings 2011 International Conference on Mechatronic Science, Electric Engineering and Computer, 2011, pp.126-129.
- [2] Kilby, J., Mawston, G., & Hosseini, H. G. . "Analysis of surface electromyography signals using continuous wavelet transform for feature extraction". IET Conference Publications, vol 23, 2006.
- [3] Ngaopitakkul, A., & Kunakorn, A. "Internal fault classification in transformer windings using combination of discrete wavelet transforms and back-propagation neural networks". International Journal of Control, Automation and Systems, vol 4, 2006, pp.365-371.
- [4] Wang, L., Wang, X., & Wei, M. "Motor bearing fault diagnosis based on wavelet packet decomposition of instantaneous power". 2010 International Conference on Computer Design and Applications, vol 3, 2010, pp.3457-3459.
- [5] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shin, H. H., Zheng, Q. "The empirical mode decomposition method and the Hilbert spectrum for nonlinear and non-stationary time series analysis". Proceedings of the Royal Society of London, Series A, vol 454, 1998, pp.903-995.
- [6] Dorostghol, A., & Dorfeshan, M. "Intelligent fault diagnosis via EMD method". Journal of Applied Sciences, vol 12, 2012, pp.1960-1965.
- [7] Tavakkoli, F., & Teshnehlab, M. "A ball bearing fault diagnosis method based on wavelet and EMD energy entropy mean". 2007 International Conference on Intelligent and Advanced Systems, 2007, pp.1210-1212.
- [8] Xu, F., & Liu, Y. F.. "Feature extraction and pattern recognition of acoustic emission signals generated from plywood damage based on EMD and neural network". Journal of Vibration and Shock, vol 31, 2012, pp.30-35.
- [9] Gu, F. C., Chang, H. C., Chen, H. C., Kuo, C. C., & Hsu, C. H.. "Partial discharge pattern recognition using Hilbert-Huang Transform in acoustic signal analysis". Proceedings of the 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition, vol 1, 2011, pp.57-61.
- [10] Guo, Z. H., Zhao, W. G., Lu, H. Y., & Wang, J. Z.. "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model". Renewable Energy, vol 37, 2012, pp.241-249.
- [11] Sun, X., Qiao, S. F., & Xie, J. R.. "The study of precipitation forecast model on EMD-RBF neural network-a case study on northeast china". Applied Mechanics and Materials, vol 641-642, 2014, pp.119-122.

- [12] Liu, J. W., Wu, X. Z., & Liu, X. X. "Time-frequency characteristic and recognition technology of acoustic emission generated from rock brittle fracture". *Applied Mechanics and Materials*, vol 405-408, 2013, pp.116-119.
- [13] Wang, Q., Zhou, Y. L., Cui, Y. F. & Sun, B. "Applied study of EMD and neural networks on flow regime identification for gas-liquid two-phase flow". *Journal of Engineering Thermophysics*, vol 28, 2007, pp.442-444.
- [14] Wang, J. J., Zhang, W. Y., Li, Y. N., Wang, J. Z., & Dang, Z. L. "Forecasting wind speed using empirical mode decomposition and Elman neural network". *Applied Soft Computing Journal*, vol 23, 2014, pp.452-459.
- [15] Jaouher, B. A., Nader, F., Lotfi, S., Brigitte, C. M., & Farhat, F. "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals". *Applied Acoustics*, vol 89, 2015, pp.16-27.

Session 14

End User Applications and Maintenance in Industry 3

Session Chair Dr. Lei Mao

Research of Optimal Placement of Sensor for Wind Turbine Based on Immune Algorithm

(Z. Song, H. Li, H. Zhang, D. Zhen, Z. Shi)

A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings

(M. Hemmer, T. I. Wang, K. g. Robbersmyr)

Design and Development of a Multipurpose Test Rig for Intelligent Lip Seals

(W. Sinzara, H. Brooks, I. Sherrington, A. Onsy, E. H, Smith M. J. Yanez, L. Marquez)

An Intelligent Maintenance System for Driverless Cars

(J. Philip, A. Onsy, M. Varley)

Research of Optimal Placement of Sensor for Wind Turbine Based on Immune Algorithm

Zhongyue Song¹, Haiyang Li¹, Hao Zhang¹, Dong Zhen^{1*}, Zhanqun Shi¹

¹ School of Mechanical Engineering, Hebei University of Technology, Tianjin, China

ABSTRACT

Signal acquisition is a prerequisite for fault analysis and diagnosis, but also the necessary condition in the fault diagnosis of wind turbine gearbox. At present, the most mature method is vibration signal analysis method, however, because the internal structure of the gearbox can not be destroyed, it can only lead to vibration sensor installed on the external wall of the gearbox. The gearbox arrangement of sensors will directly influence the quality of diagnostic accuracy, using immune algorithm to solve the problem of sensor arrangement in fault diagnosis of wind turbine, construct the gearbox vibration source coordinates, calculate the optimal transmission path of vibration transmission. To determine the best measuring point and provide an effective method for sensor arrangement.

Keywords: Immune algorithm, Points for sensor layout, Layout optimization, Wind turbines

Corresponding author: d.zhen@hebut.edu.cn

1. INTRODUCTION

In recent years, with the increasing of the total installed capacity of wind turbine, the research on the fault diagnosis of wind turbine is also increasing, the problem of how to optimal layout to the vibration sensor is also becomes a hotspot. It is a common method to optimize the measuring points of the gearbox sensor with intelligent algorithm. The immune algorithm is applied to the natural phenomena of immune mechanism in the field of Engineering. It has the advantages of parallel processing, strong robustness and so on^[1]. In the field of engineering, applying the immune algorithm to solve the problem of measuring wind the least turbine gearbox sensor to get most of the data, at the same time it will play to improve the diagnosis efficiency, which will greatly reduce the cost of diagnosis, and it has high engineering significance.

Zheng Tao of North China Electric Power University used the immune algorithm to determine the fault information of the low-voltage side of transformer substation switch, the simulation results show that the immune algorithm can effectively locate the fault of distribution network, which verified the validity and reliability of ^[2]; Yin Zhonggang of Xi'an University of technology used the immune algorithm to the internal model control of induction motor, the problem of filter time constant leads to IMC system dynamic performance was solved, and the stability of the control strategy was proved^[3]. Zhao Yu of LanZhou JiaoTong University applied the immune algorithm to the health monitoring of the bridge, and the characteristics of fast convergence speed and high diagnostic accuracy was proved^[4].

2. PRINCIPLE OF IMMUNE ALGORITHM

Immune algorithm is an important branch of intelligent algorithm, the immune system is the mechanism from biological has immune recognition, immune tolerance, immune response, immune regulation and immune memory. According to the theory of natural evolution, immune algorithm simulates the function of biological immune system. First, the system determines whether the antigen invasion, If there is an invasion of antigen, then activate memory cells to produce antibodies. Screening some antibody to update the antibody library from optimal antibody, a new antibody group builed; To calculate the affinity between antibody and antigen, The antibodies with higher fitness were differentiated into memory cells to replace the memory cells with relatively poor fitness^[5]. Promote antibodies with high fitness, inhibit antibodies with low fitness, and then produce a new antibody replacement groups. The steps of immune algorithm calculation are shown in figure 1:

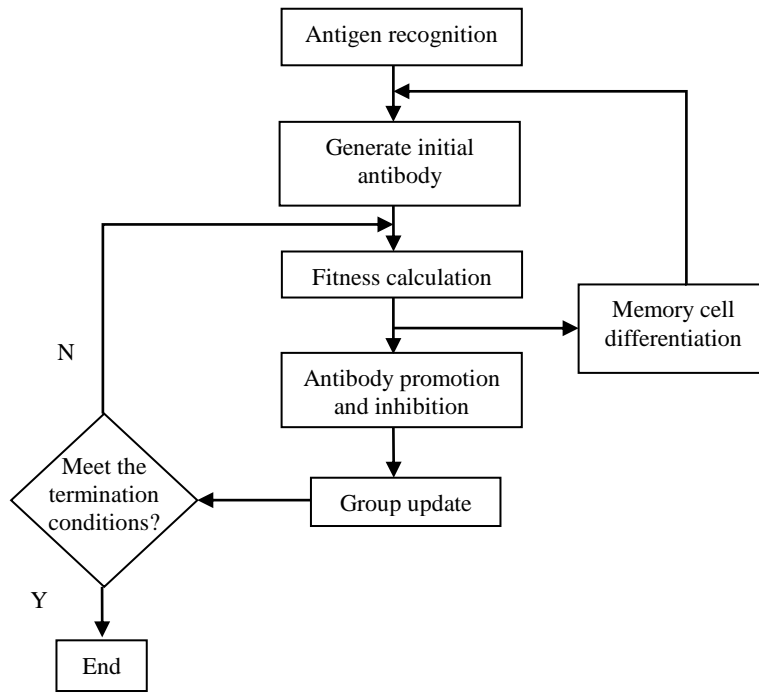


Figure 1. Immune algorithm flow.

According to the above process, the two most important links are fitness calculation and the calculation of variance before group update; The fitness calculation can largely determine the accuracy of the algorithm, and the mutation operator determines the stability and convergence of the algorithm to a great extent. The use of information entropy fitness calculation method, using G to represent a group, including N strings with length of L , M is a character set contains n characters.

$$G = \{X = x_1 x_2 \dots x_l, x_i \in M, 1 \leq i \leq l\}$$

$$H_j(G, N) = \sum_{i=1}^m -p_{ij} \log p_{ij}$$

The average information entropy of the population G is:

$$H(G, N) = \frac{1}{l} \sum_{j=1}^l H_j(G, N)$$

When the M is a binary character set, the adaptation of antibody ab_{ik} and ab_{jk} is:

$$aff(ab_i, ab_j) = H(G, 2), G = \{ab_{i,k}, ab_{j,k}\}$$

Using polynomial mutation operator to carry out mutation processing ^[6], the formula is:

$$v'_k = v_k + \delta \times (u_k - l_k)$$

In this formula:

$$\delta = \begin{cases} \left[2u + (1-2u)(1-\delta_1)^{\eta_m+1} \right]^{\frac{1}{\eta_m+1}} - 1 & u \leq 0.5 \\ 1 - \left[2(1-u) + 2(u-0.5)(1-\delta_2)^{\eta_m+1} \right]^{\frac{1}{\eta_m+1}} & u > 0.5 \end{cases}$$

In this formula:

$$\begin{aligned} \delta_1 &= (v_k - l_k) / (u_k - l_k) \\ \delta_2 &= (u_k - v_k) / (u_k - l_k) \end{aligned}$$

$$\eta_m \text{ is distribution index}$$

$$u \in [0,1]$$

Through constant iteration, finally, the solution space is the solution of the optimal sensor placement.

3. GEARBOX MODEL AND VIBRATION COORDINATE SYSTEM

The wind turbine transmission system is located in the cabin in the high altitude, so it can not be too heavy, and can not occupy a larger space, The planetary gear has the characteristics of small size, large capacity and stable operation, so the planetary gear generally used in the transmission system of wind turbine. The typical gearbox of wind turbine usually with one planetary gear transmission and two parallel shaft transmission, the gear parameters of wind turbine gearbox is shown in table 1.

Table 1: The parameter of gear

| | The number of teeth | Modulus | diameter (mm) |
|------------------------------|---------------------|---------|---------------|
| Sun gear | 80 | 3 | 240 |
| Planetary gear | 20 | 3 | 60 |
| Ring | 120 | 3 | 360 |
| Primary parallel axis | 50 | 4 | 200 |
| | 25 | 4 | 100 |
| Two parallel axes | 40 | 6 | 240 |
| | 20 | 6 | 120 |

The establishment of the coordinate system of the source of vibration has a great relationship with the structural form of the gearbox. The gear box is made up of many parts, and the position of the contact between each component is the point where there maybe a possibility of failure. The main factors that affect the probability of failure are the machining process of the part itself, the additional stress generated by the assembly error. Now the fault probability of each point is not to do the research, first according to the gearbox structure to determine the location of the fault, the gearbox structure is shown in figure 2:

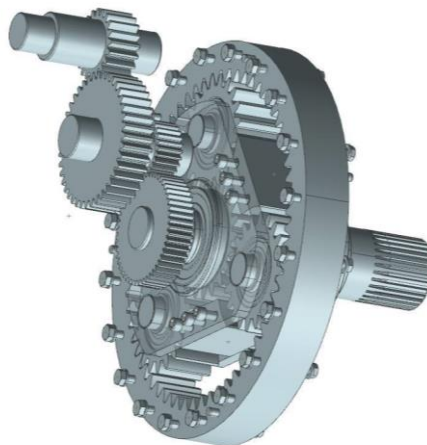


Figure 2. The model of gearbox

The coordinate system of gearbox is established according to the principle of vibration signal, the propagation of vibration signal in gear box depends on the elastic force, and the elastic force is the result of the interaction between the components, the motion of a particle drives the motion of a nearby particle, so that the wave is transmitted in turn. The transmission path can be connected by each contact point which also can be a turning point. Each gear meshing point are likely to become the source of vibration. According

to various sources of gearbox transmission path to establish the vibration source coordinates which as shown in figure3:

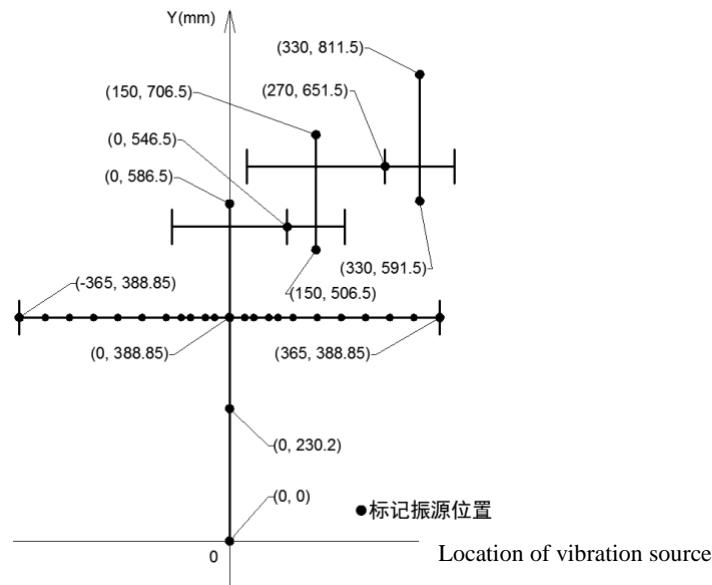


Figure 3. Vibration coordinate system of gearbox.

Because of the rotation axis of the planetary gear is moving everytime in the gear box, so the position of the planetary gear in the gear box is constantly changing, but the ring gear is fixed, the failure of any part of the are likely in the process of planetary gear meshing with the gear rim, so, the planetary gear system vibration source along the circumference of a tooth in accordance with equal probability distribution, The position of vibration source was used for immune algorithm to determine the best sensor point.

4. DETERMINE THE BEST SENSOR POINT

Usually, the sensor is installed on the wall thickness or the installation is more convenient, so we choose a set of methods to determine the initial solution of the immune algorithm; After several iterations, it converges to the optimal solution.

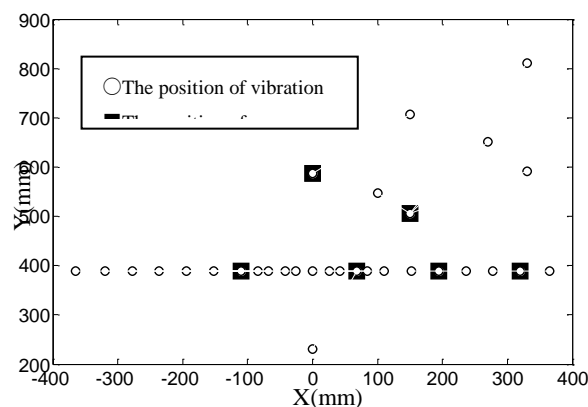


Figure 4. layout of the best sensor points

The calculation results show that the sensor measuring point arrangement is not uniformly distributed along the circumferential direction. In order to accurately detect the failure of the parallel shaft gear, some sensors are needed to be installed in the rear shell. The optimal sensor point layout is shown in figure 5.



Figure 5. The best measuring point of the sensor in gearbox

5. SUMMARY

The immune algorithm is used to optimize the sensor measuring points, a gearbox transmission model and the vibration sources distribution coordinates were built and analyzed. Through analyzing the iterative process, the optimization of the sensor arrangement was varified, then the optimal layout of the sensor can be determined for the experiments in order to obtain effective measured signals with high quality.

ACKNOWLEDGMENTS

This research was financially supported by the National Natural Science Foundation of China (Nos. 51605133), the Tianjin Natural Science Foundation of China (Grant nos. 14JCYBJC42100), and the international science and technology cooperation project of Hebei province (Grant nos. 17394303D).

REFERENCE

- [1] Zhou Juanjuan, Zhang Guangming, Yuan Yuhao. Research on fault self healing system of wind turbine based on immune [J]. mechanical design and manufacture, 2012, (3): 136-138.
- [2] Zheng Tao, Pan Yumei., Sun Jie, Jie. Study on fault location of distribution network based on immune algorithm [J]. power system protection and control, 2014, 42 (1):77-83.
- [3] Yin Zhonggang, Niu Jianbo, Zhong Yanru. An internal model control strategy of induction motor based on immune algorithm [J]. proceedings of the Chinese society of electrical engineering, 2013,33 (24): 97-105.
- [4] Zhao Yu, Yin, Peng Zhenrui. Immune algorithm in the optimization of monitoring sensor bridge application in [J]. configuration of LanZhou JiaoTong University, 2013, 32 (4): 39-43.
- [5] Wen Shihua, Zheng Jinhua, Li Miqing. Compared with the research of [J]. computer engineering and application, the mutation operator of multi objective evolutionary algorithm 2009, 45 (2): 74-78.
- [6] ZHOU Xia, SHEN Jiong. An immune based multiobjective GEP algorithm for identifying polynomial NARMAX model [J]. Control and Decision,2009,(6):1009-1015.

A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings

M. Hemmer¹, T.I. Waag², K.G. Robbersmyr¹

¹ Faculty of Engineering and Science, University of Agder, Jon Lilletuns Vei 9, N-4879 Grimstad, Norway

² Teknova AS, Tordenskjolds gate 9, N-4612 Kristiansand, Norway

ABSTRACT

Rolling element bearings (REBs) are key components in most rotating machinery. Large, slow-rotating REBs found in heavy industrial applications like offshore drilling equipment, steel- and paper mills and wind turbines are the topic of this paper. In such applications, bearings are normally non-redundant components, meaning bearing failure will cause system downtime. Execution of unplanned, on-site maintenance may be costly, time-consuming and difficult or even impossible. Implementation of condition-based maintenance strategies is a means to reduce total lifecycle costs by improving utilization of component lifetime while maintaining system availability. Condition monitoring systems capable of early, reliable detection and diagnosis of incipient faults is necessary for the planning of maintenance actions in due time. In this paper, novel and established condition monitoring methods are surveyed for this purpose. Prominent challenges are speed variations, non-stationary behavior, and low signal-to-noise ratio. Advanced signal processing methods, including order tracking and resampling from time to angular domain, higher order statistics, and cyclic spectral analysis are presented. Methods for data acquisition and maintenance decision making are also discussed. A discussion of the surveyed methods and suggestions for future research concludes the paper.

Keywords: Slow-rotating machinery, Condition Monitoring, Advanced Signal Processing, Large Roller-element Bearings, Diagnosis

Corresponding author: Martin Hemmer (email: martin.hemmer@uia.no)

1. INTRODUCTION

Roller Element Bearing (REB)s are essential mechanical components, used in virtually all types of rotating machinery. The range of types, variants and sizes match the diversity of applications. This paper aims to provide insight into condition monitoring (CM) methods suitable for large, slow-rotating bearings, typically found in paper and steel mills, offshore drilling equipment, wind turbines and similar heavy industries. It is difficult to define strict limits with regards to size and speed. However, other similarities can be defined. Replacement and maintenance is expensive, time-consuming, and in many cases not possible on site. Combined with operational non-redundancy, this motivates CM for increased control of machine health. The ability to utilize more of the component lifetime while reducing the risk of unexpected failure potentially reduces lifecycle costs.

A survey presented at the Noble Analyst Day 2012 mapped causes of downtime on drilling rigs in the period 2011-December 2012 [1]. Looking at downtime by equipment on all rig types, top drive failure is the second largest contributor, with a total of 13 %. Further analysis shows that bearing failure is the main overall cause of downtime in top drives, despite variations between manufacturers and types. In top drives, a large tapered roller thrust bearing supports the weight of the drill string. With an outer diameter of up to 750mm and a rotational speed of 240 rpm (4 Hz), top drive main bearings qualify as both large and slow-rotating. The statistics show a potential for improved CM of large, slow-rotating bearings in the offshore industry.

2. CONDITION BASED MAINTENANCE STRATEGIES

Most systems require maintenance to a certain extent. The approach to maintenance is influenced by factors such as consequences of failure, maintenance cost, and failure rates. Development in technology has enabled more advanced maintenance strategies beyond corrective and preventive maintenance strategies. Lee et.al [2] reviewed the field of Prognostics and Health Management (PHM) for rotating machinery, presenting a generalized methodology for selection and implementation of a maintenance strategy. A maintenance transformation map is proposed as a guideline for maintenance strategy selection based on system complexity and uncertainty. Typically, condition monitoring of bearings falls in the Condition-Based Maintenance (CBM) category. However, as bearings get larger, maintenance is more complicated. External factors like long spare part lead times, complicated maintenance procedures and limited maintenance opportunities add complexity, justifying strategies like PHM. Common for both CBM and PHM is the need for reliable health assessments for the equipment. Jardine et.al [3] reviews generalized diagnostics and prognostics methods for successful implementation of Condition Based Maintenance (CBM), identifying data acquisition, data processing and maintenance decision-making as the three main steps. The influence of bearing size and speed on these steps is discussed in the next section.

3. CHALLENGES OF LARGE, SLOW BEARINGS

Large and slow-rotating bearings pose challenges for conventional condition monitoring methods. Compared to smaller bearings operating at higher speeds, prominent challenges include low vibration energy, sensitivity to speed fluctuations and a need for accurate localization in the frequency domain to isolate fault frequencies.

It is commonly accepted that discrete faults in bearings cause impulse-like impacts when the fault interacts with another rolling surface. This impact triggers a transient response at resonance frequencies in the bearing, surrounding structure and transducer. As acceleration is the second derivative of displacement, a reduction in rotational speed leads to significant reduction in acceleration levels. For slow-rotating bearings, fundamental fault frequencies will also be relatively close in the frequency domain, increasing the risk of interference. Bechhoefer recommends a minimum of 10, preferably 30, frequency bins between fault frequencies [4]. Frequency resolution is the inverse of acquisition time, which makes the measurement more prone to capture speed fluctuations, leading to smearing of the frequency spectrum. Larger size also means increased distance between fault and transducer. All these factors contribute to a lower signal to noise ratio (SNR).

4. MODELLING BEARING FAULTS

The periodic nature of impacts can be modeled mathematically. Time between impacts is governed by a combination of shaft speed, bearing geometry and localization of the fault. For REBs, faults can be associated with the characteristic fault frequencies of the bearing components, Ball Pass Frequency Inner race (BPFI), Ball Pass Frequency Outer race (BPFO), Cage Pass Frequency (CPF) and Ball Spin Frequency (BSF) [5]. Normalizing by shaft frequency transforms frequencies to shaft order domain for easier comparison across operating speeds.

Characteristic fault frequencies assume ideal operating condition, including perfect rolling motion between rolling elements and races. In reality, rolling elements experience some random slip, causing variation in the time between impacts. Additionally, the impact response amplitude can be periodically modulated with smaller random variations. Antoni includes this randomness in a more realistic model for a bearing vibration signal, given in equation (1) [6]. The vibration signal is $x(t)$, where $h(t)$ is the response to a single impact, $q(t) = q(t + P)$ is periodic modulation caused by load distribution of period P , and T represents the time between the arrival of two consecutive impacts. The random jitter in arrival time and amplitude is handled by τ_i and A_i respectively.

$$x(t) = \sum_{i=-\infty}^{i=\infty} h(t - iT - \tau_i) q(iT) A_i + n(t) \quad (1)$$

This randomness in arrival times causes smearing in the frequency spectrum, but allows separation of the bearing signal from deterministic frequency components from gears and shafts [5].

4.1. Cyclostationarity

Processes that shows cyclic behavior is said to exhibit cyclostationarity [6]. Cyclostationary theory provides a generalized framework for describing a wide range of stationary and non-stationary processes [7]. In the context of cyclostationarity, a periodic component at a frequency α is referred to as the cyclic frequency. The period of α is termed cycle. These terms are used to avoid confusion with spectral frequency f and its period T . Figure 1 shows a time signal, highlighting the difference by indicating the cycle of α and period of f .

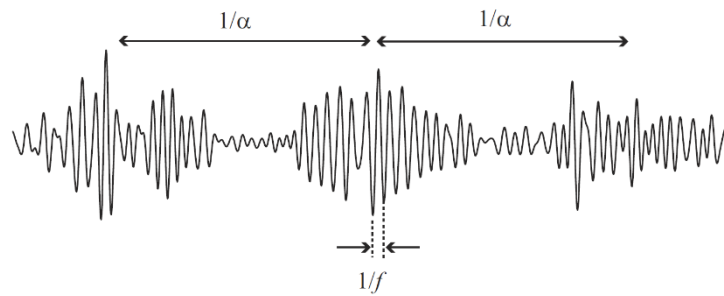


Figure 1. The difference between cyclic frequency and spectral frequency [7].

A signal can exhibit cyclostationarity at different orders. As an example, a periodic signal masked with additive white noise will have a periodic mean value and thus exhibit first-order cyclostationarity. Consider a signal of amplitude-modulated white noise only. As the mean value is constant, no periodic first-order components exist. Squaring the signal, a second-order transformation, reveals periodic components and consequently second-order cyclostationarity in the signal. A second order transformation is normally enough to reveal bearing diagnostics information. Interested readers can consult the works of Randall [5], [8], [9] and Antoni [6], [7], [10] in particular for further information on the topic and its applications. Cyclic spectral analysis, based on cyclostationary theory, is introduced in section 5.5.

5. CONDITION MONITORING METHODS

This section is divided in three, discussing methods for data acquisition, data processing and maintenance decision-making; identified by Jardine et.al [11] as the three main steps in CBM. Here, signal enhancement is included as data processing.

5.1. Data Acquisition

Choosing a measurement technique capable of observing the symptoms of failure is critical. Tandon and Choudhury [12] identifies four main categories for bearing fault detection methods; vibration measurements, acoustic measurements, lubrication analysis and temperature measurements.

5.1.1. Vibration Measurements

Vibration monitoring using accelerometers is widely used in the industry, and has been researched actively since the 1980s [12]. Traditional vibration analysis faces some challenges when applied to large, slow bearings, as discussed in section 4.1. Displacement measurements can also be used for monitoring vibration. Measuring displacement directly instead of acceleration makes it suitable for slow applications with low acceleration levels. Shakya et.al [13] investigated the use of proximity probe as a standalone

method for bearing fault detection and in combination with an accelerometer, and showed improved detectability for inner race defects.

5.1.2. Acoustic Measurements

Acoustic measurements refer to vibrations from 20 kHz and upwards, including both ultrasonic and Acoustic Emission (AE) measurements. An increase in AE activity could be an early indication of oil degradation. AE activity in bearings can be related to metal-to-metal contact, indicating a broken oil film. Yoshioka and Fujiwara [14], [15] showed in early research that AE could detect faults before vibration methods. Further research successfully used AE for detection of subsurface cracks [16], which Tan [17] concluded could be useful for detection of pitting. Chacon et. al [18] presented a method for incipient fault detection in REBs using AE measurements for envelope analysis. Jamaludin and Mba review monitoring of extremely slow REBs [19], [20], using AE measurements to detect faults at very low speeds.

5.1.3. Lubrication Analysis

Many bearing failure modes can be related to insufficient lubrication; fatigue, wear, corrosion, deformation, and fracture. Examples of condition indicators from lubrication analysis are accumulated particle mass, water content, viscosity, conductivity and debris analysis. Dempsey [21] compared oil debris analysis and vibration based CIs for detection of pitting. All CIs increased when pitting occurred. However the article highlighted the need for improved threshold setting and combination of CIs for improved reliability. Bechhoefer et.al [22] compared lubricant, vibration and temperature data from wind turbine bearings. Lubrication analysis indicated a fault in one damaged bearing but also gave one false alarm. In combination with vibration analysis better accuracy was achieved.

5.1.4. Temperature Measurements

The use of temperature as a CI alone is not likely sufficient, especially in offshore equipment where ambient temperature will vary. Load variations will also affect bearing temperature, and lower rotational speed will generate less heat than in high-speed bearings. This combination makes temperature measurements less suitable for condition monitoring. In a comparative study of vibration, lubricant analysis and temperature for condition monitoring [22], temperature failed to indicate failure on a large, slow-rotating wind turbine bearing.

5.2. Signal Enhancement

In cases where information carried in the signal is severely masked, pre-processing techniques can be applied to separate, enhance or in other ways improve the signal of interest. The methods do not provide any diagnostics information on their own, but facilitate the use of other CM methods.

5.2.1. Correcting for Speed Variation

Variations in shaft rate during sampling will distribute frequency content across more bins in the spectra. As discussed, characteristic fault frequencies are close for large, slow-rotating bearings, which makes low-speed applications sensitive to variations in shaft speed. Order tracking corrects for shaft speed variations by using shaft angle as a reference instead of time. In addition to the original measurement, shaft angle is recorded simultaneously. Measured data is then resampled to angular domain, which effectively manipulates the sampling frequency. Bechhoefer et. al [4] recorded speed variations in the range of 2% on a wind turbine, which caused smearing of frequency content. Order tracking effectively removed these variations, resulting in a sharper spectrum. Resampling vibration data using improved the detection and diagnosis [23]–[25].

5.2.2. Isolating the Bearing Signal

Shafts and gears can interfere heavily with the bearing signature, particularly in slow-rotating applications where the SNR of fault signatures is low. Linear prediction, adaptive and self-adaptive noise cancellation, Discrete/random separation and Time Synchronous Averaging (TSA) for use in bearing diagnostics are

presented by Randall and Antoni in [5]. Borghesani et.al [26] demonstrated a cepstrum based pre-whitening method for extraction of the bearing signature. Common for all methods is that the bearing signature can be isolated from discrete frequency components by exploiting the randomness of bearing vibration as opposed to deterministic gear and shaft signatures.

5.3. Established CM Methods

In the context of bearing condition monitoring, analysis of time waveform data from vibration, acoustic and ultrasonic is the established industry standard, with notable work by Tandon, Nakra and Choudhury [12], [27], Kim et.al [28], Ho [29] and Randall [8], [30]. Fourier Analysis is fundamental CM, particularly using the Discrete Fourier Transform (DFT). However, diagnostics information often lies in the periodic modulation of a given carrier frequency. Thus, Fourier analysis of the raw signal alone might not be able to detect or diagnose faults. Envelope analysis is perhaps the best example, widely regarded as a benchmark for bearing fault detection [5]. Here, Fourier analysis is preceded by bandpass filtering around an assumed carrier frequency before the signal envelope is calculated. A challenge is to choose the correct bandpass filter. In digital signal processing, envelope extraction of a signal is often done by taking the absolute value of its Hilbert transform [5]. Then, Fourier analysis is performed on the envelope to reveal frequency component corresponding to the fault frequencies from section 4. Envelope analysis and applications have been thoroughly examined in [8], [12], [27], [29]–[31].

Other methods can be applied directly in time domain, such as Root-Mean-Square (RMS) and Crest Factor (CF). Kim et. Al [28] compared vibration and ultrasonic measurements for bearing fault detection across a range of low speeds. RMS was shown to decrease almost linearly with shaft speed. However, healthy bearings have shown big variance in RMS values, indicating RMS change is a better CI than predefined threshold values. CF, the ratio of peak amplitude to RMS value, will increase immediately when a fault first appears. Williams et.al [32] recorded the CF in bearing run-to-failure experiments, and reported an increase followed by a decrease as the fault developed. This indicates fault detection capabilities, but limitations as a trendable parameter for diagnostics purposes.

5.4. Advanced CM Methods

Traditional CM methods are often insufficient for reliable fault detection in large, slow-rotating bearings. This section presents a selection of advanced and novel CM methods available to overcome the challenges.

5.4.1. Higher Statistical Moments

For nominal bearings, the acceleration Probability Density Function (PDF) can be assumed to have a Gaussian distribution. Thus, any changes in the shape of the PDF can indicate failure [12]. Statistical moments of first and second order, mean (μ) and standard deviation (σ) respectively, are well known. For CM purposes, moments of a higher order k , calculated as in equation (2). Mainly skewness ($k=3$) and kurtosis ($k=4$), are used as CIs.

$$\frac{1}{N} \sum_{n=1}^N \left(\frac{x[n] - \mu}{\sigma} \right)^k \quad (2)$$

Skewness describes the asymmetry of a distribution, i.e. the relative energy above and below the mean. Nguyen et.al [33] identified skewness as one of three optimal features for reliable fault detection in low-speed bearings, but skewness is not consistently reported as a reliable CI. Kurtosis is as a measure of tailedness, i.e. the presence of tail extremities in a dataset [34]. A Gaussian distribution always has a kurtosis of 3. High amplitude accelerations from impacts yield a heavy-tailed distribution and high kurtosis. This makes kurtosis suitable as a standalone CI, requiring no prior knowledge to quantify the condition.

5.4.2. Spectral Kurtosis and the Kurtogram

Spectral Kurtosis (SK) identifies non-Gaussian components in signals along with their location in the frequency spectrum. The method was proposed in 1983 by Dwyer [35]. Wang, Y et. al [36] and Wang, P. et. Al [37] published a review on the use of SK for fault detection, diagnostics, and prognostics for bearings. SK has also been shown to aid optimal selection of frequency band for envelope analysis [38].

The kurtogram was proposed by Antoni in [39], mapping SK as a function of center frequency and filter bandwidth. Wang, P. et. al [37] similarly utilized SK for frequency band selection, but proposed an enhanced kurtogram based on kurtosis of the power spectrum.

5.4.3. Wavelets

A wavelet is a waveform with a limited duration that integrates to zero and can be scaled and shifted in time. The Wavelet Transform (WT) provides a time-scale representation of the signal, where scale is qualitatively comparable to frequency. An important advantage is the good time resolution at high frequencies and high frequency resolution at low frequencies. Klepka presented a wavelet-based demodulation technique [40], which combined the use of the continuous and discrete-time WT for filtering, envelope estimation and fault detection on synthetic bearing data. Gelman et. al [41] proposed an improved method, using SK for optimal selection of frequency band while maintaining the advantages of wavelet demodulation compared to Fourier analysis.

5.4.4. Empirical Mode Decomposition

Empirical Mode Decomposition (EMD), also known as the Hilbert-Huang Transform, obtains instantaneous frequency information of an oscillatory signal by separating it into several Intrinsic Mode Functions (IMFs) which can be amplitude- and frequency-modulated non-linearly. EMD is the data-driven and adaptive, as IMFs are based on the sampled signal only. Lei et.al [42] reviews the application of EMD to fault diagnosis of rotating machinery. Žvokelj et. al [43] demonstrated fault detection on large, slow rotating bearings using EMD and Principal Component Analysis (PCA).

5.4.5. Cepstrum Analysis

The (real) cepstrum, defined in equation (3), identifies repeating “echoes” of a signal, which can be used for detection of periodic signatures. Bechhoefer et.al [4] tested cepstrum analysis for fault detection on wind turbine main bearings, and observed indications of an outer race fault. Cepstrum RMS and kurtosis were tested as possible CIs, but were not able to give actionable results alone. Further study was recommended.

$$\text{cepstrum} = IFT\{\ln|FT\{x\}|\} \quad (3)$$

5.5. Cyclic Spectral Analysis

Cyclic spectral analysis relies on the concept of cyclostationarity from section 4.1. Here, two approaches for detecting cyclostationarity in signals will be introduced. The first extracts periodic components of the instantaneous power, while the second is based on the autocorrelation function. An operator $P\{\bullet\}$ is presented in [7]. The operator is implemented as an estimator, shown in equation (4), where n is the sample number and T_s is the sample period. The estimator extracts Fourier coefficients at cyclic frequencies α in set A , from a given discrete data sequence $\{\bullet\}$.

$$\hat{P}\{\bullet\} = \sum_{\alpha \in A} DFT_{\alpha} \{\bullet\} e^{j2\pi\alpha n T_s} \quad (4)$$

5.5.1. Power Decomposition and Instantaneous Autocorrelation

Cyclostationary behavior can be detected by decomposition of signal power to periodic components. Consider a signal $x[n]$ with power P_x . Estimation of mean instantaneous power $P_x[n]$ is done by applying $\hat{P}\{\bullet\}$ to the signal power $|x[n]|^2$. A Fourier series expansion of $P_x[n]$ then gives the cyclic powers P_x^{α} . The quantities mean instantaneous power spectrum and cyclic modulation spectrum are obtained by a time frequency decomposition and Fourier series expansion respectively, further elaborated in [7].

The presence of periodicity in a signal creates a correlation of spectral components at the cyclic frequency. Antoni describes in [6], [7] how the instantaneous autocorrelation function can be utilized to detect cyclostationary behavior. Given a signal $x(t)$, the instantaneous autocorrelation $R_x(t, \tau)$ is defined as applying $P\{\bullet\}$ to the symmetric autocorrelation function, shown in equation (5).

$$R_x(t, \tau) = P\{x(t + \tau/2)x(t - \tau/2)\} \quad (5)$$

The Fourier series expansion of the instantaneous autocorrelation function expressed is called the cyclic autocorrelation function $R_x^\alpha(\tau)$. Applying a Fourier transform to $R_x^\alpha(\tau)$ yields the spectral correlation density $SC_x^\alpha(f)$, a frequency-frequency representation of $x(t)$. Note that cyclic frequency α is the frequency counterpart of time, and spectral frequency f is the dual of shift τ . The spectral correlation (SC) is non-zero if a frequency component f is periodic with cyclic frequency α . These connections and the relationship to the Wigner-Ville spectrum, classical autocorrelation function, and the PSD are examined in [7].

5.6. Maintenance Decision-making

A condition monitoring system should be able to make or aid in maintenance decisions. However, a single CIs may be insufficient to provide reliable decisions. Fusion of data of different types, both on sensor and feature level can be utilized for improved diagnostics and prognostics of bearings [2]. Dempsey and Loutas [21], [44] investigates a combination of on-line oil analysis, AE and vibration as a way of improving CI performance. Bechhoefer et. al [45] presents a method for optimal threshold setting, by fusing several CIs in a Health Index (HI), which quantifies bearing damage without the need for user interpretation. The HI is constructed from the norm of n Gaussian CIs, and can be shown to form a Nakagami-distributed PDF. This method allows for setting a desired Probability of False Alarm (PFA) and normalizing the HI to be 1 when this probability is reached. The method is successfully demonstrated on data from three large, slow wind turbine bearings [4], where the faulty bearing was shown to have a HI well above one.

6. CONCLUSION

This paper presents an overview of relevant CM methods for large, slow-rotating bearings. The combined requirement of cost reduction and uptime facilitates the emergence of more advanced condition monitoring systems. Main challenges of condition monitoring of large, slow-rotating bearings can be summarized by a low energy impacts, large distance from fault to transducer, comprehensive background noise and speed variations, resulting in a low SNR. Detection capabilities of traditional CM methods, especially envelope analysis, can be improved by longer acquisition times, order tracking and separation of random and discrete components. SK aided bandpass filtering before envelope extraction further improves performance. Other data acquisition methods can also be used. AE signals carries similar diagnostics information as vibration, but in a frequency band less subjected to noise.

Another development in bearing condition monitoring is the transition from a stationarity assumption implicated by the Fourier transform to a more realistic, non-stationary or cyclostationary approach. Time-frequency and cyclostationary analysis tools takes this into account. In cases where cyclic behavior is heavily masked in non-stationary signals, CS analysis appears to be a powerful tool.

It seems unlikely to find a single CI, data acquisition or signal processing method that solves all challenges for CM of large, slow-rotating REBs. Hence, combining CM data from different sources seems more reasonable. The concept of a PFA-controlled HI is attractive from an operator point of view, and can preferably be utilized in systems for automated fault detection and diagnostics. Finding good CIs and methods for fusing them should be a priority in future work.

ACKNOWLEDGEMENT

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, project number 237896.

REFERENCES

- [1] L. Jeffrey, "Noble 2012 Analyst & Investor Day presentation." 2012.
- [2] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, pp. 314–334, 2014.
- [3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [4] E. Bechhoefer, R. Schlanbusch, and T. I. Waag, "Techniques for Large, Slow Bearing Fault Detection," *Int. J. Progn. Heal. Manag.*, vol. 7, no. 1, pp. 1–12, 2016.
- [5] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics-A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, 2011.
- [6] J. Antoni, "Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions," *J. Sound Vib.*, vol. 304, no. 3–5, pp. 497–529, 2007.
- [7] J. Antoni, "Cyclostationarity by examples," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 987–1036, 2009.
- [8] R. B. Randall, "State of the Art in Monitoring Rotating Machinery – Part 2," *J. Sound Vib.*, vol. 38, no. 5, pp. 10–17, 2004.
- [9] R. B. Randall, J. Antoni, and S. Chobsaard, "The Relationship Between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 945–962, 2001.
- [10] J. Antoni, "Cyclic spectral analysis in practice," *Mech. Syst. Signal Process.*, vol. 21, pp. 597–630, 2007.
- [11] A. H. C. Tsang, W. K. Yeung, A. K. S. Jardine, and B. P. K. Leung, "Data management for CBM optimization Data management for CBM optimization," *J. Qual. Maint. Eng.*, vol. 12, no. 1, pp. 37–51, 2006.
- [12] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribol. Int.*, vol. 32, pp. 469–480, 1999.
- [13] P. Shakya, A. K. Darpe, and M. S. Kulkarni, "Bearing diagnosis using proximity probe and accelerometer," *Meas. J. Int. Meas. Confed.*, vol. 80, pp. 190–200, 2016.
- [14] T. Yoshioka and T. Fujiwara, "Application of acoustic emission technique to detection of rolling bearing failure," *Am. Soc. Mech. Eng.*, vol. 14, no. 1, pp. 55–76, 1984.
- [15] T. Yoshioka and T. Fujiwara, "A new acoustic emission source locating system for the study of rolling contact fatigue," *Wear*, vol. 81, no. 1, pp. 183–186, 1982.
- [16] T. Yoshioka, "Detection of rolling contact sub-surface fatigue cracks using acoustic emission technique," *Lubr. Eng.*, vol. 4, no. 1, 1993.
- [17] C. K. Tan and D. Mba, "Identification of the acoustic emission source during a comparative study on diagnosis of a spur gearbox," *Tribol. Int.*, vol. 38, no. 5, pp. 469–480, 2005.
- [18] J. L. Ferrando Chacon, V. Kappatos, W. Balachandran, and T.-H. Gan, "A novel approach for incipient defect detection in rolling bearings using acoustic emission technique," *Appl. Acoust.*, vol. 89, pp. 88–100, 2015.
- [19] N. Jamaludin and D. Mba, "Monitoring extremely slow rolling element bearings: Part I," *NDT E Int.*, vol. 35, no. 6, pp. 349–358, 2002.
- [20] N. Jamaludin and D. Mba, "Monitoring extremely slow rolling element bearings: Part II," *NDT E Int.*, vol. 35, no. 6, pp. 359–366, 2002.
- [21] P. J. Dempsey, "A Comparison of Vibration and Oil Debris Gear Damage Detection Methods Applied to Pitting Damage," 2000.
- [22] E. Bechhoefer, R. Schlanbusch, and T. I. Waag, "Fault Detection on Large Slow Bearings," in *PHME 2016*, 2016, vol. 7, pp. 1–8.
- [23] S. Priya, M. R. Ramesh, and V. Naidu, "Bearing Health Condition Monitoring: Frequency Domain Analysis Multi-sensor Data Fusion," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng. (An ISO Certif. Organ.)*, vol. 3, no. 5, pp. 260–268, 2014.
- [24] L. Renaudin, F. Bonnardot, O. Musy, J. B. Doray, and D. Rémond, "Natural roller bearing fault detection by angular measurement of true instantaneous angular speed," in *Mechanical Systems and Signal Processing*, 2010, vol. 24, no. 7, pp. 1998–2011.
- [25] E. Bechhoefer and M. Kingsley, "A review of time synchronous average algorithms," in *Annual Conference of the Prognostics and Health Management Society*, 2009, pp. 24–33.
- [26] P. Borghesani, P. Pennacchi, R. Ricci, and S. Chatterton, "Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals," *Mech. Syst. Signal Process.*, vol. 40, no. 1, pp. 38–55, 2013.
- [27] N. Tandon and B. C. Nakra, "Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings," *Tribol. Int.*, vol. 25, no. 3, pp. 205–212, 1992.
- [28] E. Y. Kim, A. C. C. Tan, J. Mathew, and B. S. Yang, "Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements," in *Australian Journal of Mech. Engineering*, 2008, vol. 5, no. 2, pp. 177–189.
- [29] D. Ho and R. B. Randall, "Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signal," *Mech. Syst. Signal Process.*, vol. 14, no. 5, pp. 763–788, 2000.
- [30] R. B. Randall, "State of the Art in Monitoring Rotating Machinery – Part 1," *J. Sound Vib.*, vol. 38, no. 5, pp. 14–21, 2004.
- [31] D. Hochmann and E. Bechhoefer, "Envelope bearing analysis: Theory and practice," in *IEEE Aerospace Conference Proceedings*, 2005, pp. 3658–3666.
- [32] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess, "Rolling Element Bearing Diagnostics In Run-To-Failure Lifetime Testing," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 979–993, 2001.
- [33] P. Nguyen, M. Kang, J. Kim, and K. Jong-Myon, "Reliable Fault Diagnosis of Low-Speed Bearing Defects Using a Genetic

- Algorithm,” in *PRICAI 2014: Trends in Artificial Intelligence 2014*, 2014, pp. 248–255.
- [34] P. H. Westfall, “Kurtosis as Peakedness, 1905–2014. R.I.P.,” *Am. Stat.*, vol. 68, no. December, pp. 191–195, 2014.
 - [35] R. Dwyer, “Detection of non-Gaussian signals by frequency domain Kurtosis estimation,” *ICASSP '83. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 8, pp. 607–610, 1983.
 - [36] Y. Wang, J. Xiang, R. Markert, and M. Liang, “Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications,” *Mech. Syst. Signal Process.*, vol. 66–67, pp. 679–698, 2016.
 - [37] D. Wang, P. W. Tse, and K. L. Tsui, “An enhanced Kurtogram method for fault diagnosis of rolling element bearings,” *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 176–199, 2013.
 - [38] E. Bechhoefer, M. Kingsley, and P. Menon, “Bearing envelope analysis window selection using spectral kurtosis techniques,” in *2011 IEEE International Conference on Prognostics and Health Management, PHM 2011 - Conference Proceedings*, 2011, pp. 1–6.
 - [39] J. Antoni and R. B. Randall, “The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines,” *Mech. Syst. Signal Process.*, vol. 20, pp. 308–331, 2006.
 - [40] A. Klepka, “Wavelet Based Signal Demodulation Technique for Bearing Fault Detection,” *Mech. Mech. Eng.*, vol. 15, no. 4, pp. 63–71, 2011.
 - [41] L. Gelman, T. H. Patel, G. Persin, B. Murray, and A. Thomson, “Novel Technology Based on the Spectral Kurtosis and Wavelet Transform for Rolling Bearing Diagnosis,” *Int. J. Progn. Heal. Manag.*, vol. 4, no. 2, pp. 1–7, 2013.
 - [42] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” *Mech. Syst. Signal Process.*, vol. 35, no. 1–2, pp. 108–126, 2013.
 - [43] M. Žvokelj, S. Zupan, and I. Prebil, “Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemble Empirical Mode Decomposition method combined with Principal Component Analysis,” *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1049–1067, 2010.
 - [44] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos, “The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery,” *Mech. Syst. Signal Process.*, vol. 25, no. 4, pp. 1339–1352, 2011.
 - [45] E. Bechhoefer and A. P. F. Bernhard, “A generalized process for optimal threshold setting in HUMS,” in *IEEE Aerospace Conference Proceedings*, 2007, pp. 1–9.

Design and Development of a Multipurpose Test Apparatus for Intelligent Lip Seals

W. Sinzara¹, H. Brooks¹, I. Sherrington¹, A. Onsy¹, E. H. Smith¹, M. J. Yanez^{1,2}, L. Marquez^{1,2}.

¹ Jost Institute for Tribotechnology, University of Central Lancashire, Preston PR1 2HE, UK.

² Tribology Group, SEPI ESIME, Instituto Politécnico Nacional, Mexico City Mexico.

ABSTRACT

This paper describes the design of a novel test apparatus for the development of intelligent lip seals. The intelligent lip seal will include sensors, actuators and intelligent control, allowing it to react optimally to changing scenarios (such a device is also called a tribotronic system). Intelligent sealing systems offer the potential for improved function and better maintainability, with benefits including extended component life, reduced friction and wear, lower power consumption and real-time duty information. Conventional test apparatus is not ideal for developing intelligent mechanical sealing components as they lack the flexibility to test a wide range of sealing challenges and have limited data collecting capabilities.

The new test apparatus includes sensors for the measurement of shaft torque, oil temperature, vibration, shaft speed and lubricant leakage. A transparent tube is used to enable visualization of the shaft-seal counterface and other phenomena associated with the shaft-seal contact. The seal housing is repositionable to allow axial offsets between the shaft and the seal. This will allow the testing and development of seals capable of mitigating leakage or reducing friction in the presence of misalignment.

The test apparatus also supports variable shaft speeds, lubricant temperatures, lubricant pressures and interchangeable shafts with varying surface roughness in order to carry out testing under a range of conditions.

Keywords: lip seal, intelligent machines, test apparatus

Corresponding author: Wilbert Sinzara (wsinzara@uclan.ac.uk)

1. INTRODUCTION

A seal is a device used to control the interchange of fluids between two regions that share a common boundary. Where the boundary comprises components in relative motion, dynamic seals are used. These include rotary mechanical seals (both radial and axial), rubber lip seals, labyrinth seals and piston rings [1]. In a study conducted by a pump manufacturer it was noted that malfunctioning seals account for up to 39% of pump failures [2]. Increasing the reliability and efficiency of motors, pumps, gearboxes and similar equipment is a major challenge for all OEMs. If friction at the seal-shaft interface is not properly managed, a substantial amount of heat can be generated. Not only is this heat a waste of energy which affects operating costs, it can accelerate deterioration of the seal face, and may add to cooling loads. This paper is concerned with mechanical lip seals. These are used in all industries, including oil, gas, water, power, transport, manufacturing, and aerospace. In 2015, the mechanical seal market earned revenues of \$3.6 billion according to global market research company Frost and Sullivan and the market is forecast to grow at 9% compound annual growth rate between 2017 and 2021[3].

As is the case with all tribological elements, sealing systems are dependent on disciplines as diverse as fluid mechanics, lubrication, materials science, heat transfer, phase changes, friction and design. Physical processes at the seal interface are determined by molecular interaction, surface finish, heat conduction and transfer, oil film thickness, relative velocity and material properties. The complexity that arises as a result of the interaction of all these phenomena makes it difficult to develop a simple analytical model of seal

behaviour. Due to this complexity, it is necessary to develop test apparatus to assist in studying these interacting phenomena.

This paper discusses the design and development of a novel test apparatus for the parametric testing of intelligent (tribotronic) lip seals. The test apparatus will allow the study of sealing friction and function, over the seal life, for a range of shaft speeds, surface roughness, axial offsets, lubricant temperatures and lubricant pressures. A camera will also allow visual inspection of the seal-shaft interface.

1.1. Tribotronic Lip Seals

A typical profile of a mechanical lip seal is shown in figure 1. The garter spring provides constant circumferential pressure on the contact face. The problem with applying radial sealing force in this manner is that the pressure of the spring is applied uniformly around the entire circumference and changes over time due to creep, wear and temperature change. Additionally, different sealing pressures are required if the lubricant's temperature and pressure change, and the spring tension is also unlikely to be optimal if the shaft is misaligned.

Seals are passive components, and many of the main operational issues connected with them are associated with the control of friction, lubrication and wear. Consequently, two of the main aims of their design are the establishment of long life and efficient function in terms of both sealing ability and sliding efficiency. Each device is designed or selected to operate effectively up to a specific limiting speed and load for a given period. Tribotronics is a novel concept in which sensors and actuators are employed in tribological components or systems which would traditionally be passive, to dynamically modify their performance during operation [4, 5]. The authors aim to develop seals that actively monitor and control themselves to enhance their own health/life and/or performance. In addition, they will be "connected", that is able to feedback information about operating duty, performance and condition. This includes incorporating opportunities to control and manage the seal and any associated systems via the "internet of things". This can be done either for maintenance purposes, for machine learning reasons (to enhance the performance of other devices in a connected system) or for other aspects of functional enhancement, such as friction reduction. Tribotronic seals have the potential to mitigate the aforementioned problems by intelligently adapting to the changing conditions. A typical process control loop schematic for a tribotronic seal is shown in figure 2.

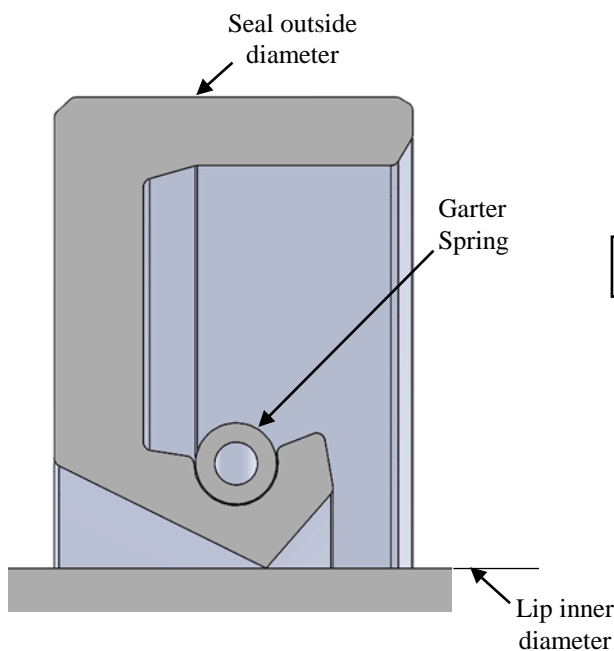


Figure 1. Conventional lip seal profile

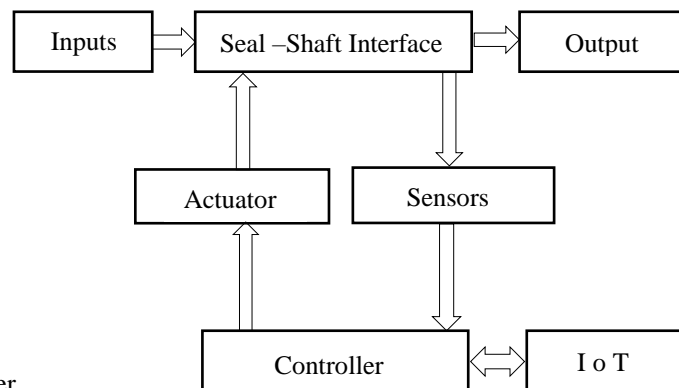


Figure 2. Control loop for tribotronic seal

2. LITERATURE SURVEY

2.1. Lip Seals

The first seal materials were leather and wood. The first patents for (leather) lip seals started appearing in the early 1900s and the first patent to incorporate a garter spring, to apply a constant radial force on the shaft counter face, was registered by Freudenberg in 1932 [15]. Up to the present day, the basic features of the lip seal have not changed much in terms of operating principle. The most notable changes are the adoption of polymeric materials, and the incorporation of multiple lips for both liquid-side and air-side sealing. In some applications where seal failure is critical and unpredictable, engineers resort to redundant sealing, whereby more than one seal is installed in the same seal cavity. From the beginning of the 2000's, OEMs started introducing sensing in seal elements. These are capable of measuring temperature (using thermocouples), leakage (using capacitance based methods) and lubricant condition [16-18].

Conventional lip seals are passive elements because they neither sense their condition and environment nor use actuators to adapt during operation. These passive systems have been in use for 100 years or more. In 1957 Jagger posited and proved that for a properly functioning seal the load imposed by the seal lip on the shaft is carried by the oil film at the interface [6]. This hypothesis was advanced in the 60's by Johnson and Gabelli [7, 8]. By the late 1980's and 90's the hypothesis that the lubricant is kept inside the oil side by a "reverse pumping" mechanism took hold [9-11]. Complex algorithms of the elastohydrodynamic "reverse pumping" models were developed to support and explain this phenomenon [12, 13]. Hidrovo and Hart of MIT designed and developed a charged couple device (CCD) camera technique to visualise events at the seal interface [14]. They successfully used this technique to study seal failure due to dirt ingress in the seal face.

In order to investigate each of the above hypotheses it was necessary to develop various test apparatus to validate or verify the underlying assumed theories. A few selected test rigs are noted in section 2.2.

2.2. Lip Seal Test Apparatus

The British Standard BS7780-4 sets out the procedures for dynamic seal testing and also includes generic test bench designs for dynamic and low-temperature testing. The test procedure is suitable for performance testing of conventional lip seal products, with a particular emphasis on endurance testing. Since the standard sets a basic guideline, OEMs and research laboratories develop in-house test procedures to develop seals and validate their own data. Specific apparatus are developed to measure the individual parameters, usually resulting in multiple measuring devices. One such single-parameter measuring device used a charged couple device camera to observe ingress of dirt using a laser-induced fluorescent dye [14]. In a similar investigation, Ogata et al used a CCD video camera to study the behaviour of circumferential micro undulations with axial grooves on the lip surface [19]. The test rig uses optics techniques to measure the rate of a decaying light beam shone through the lubricant, to predict the oil film thickness along the ridges. Stakenborg developed a test bench which used a square bundle of step-index multi-mode glass fibres inserted into a hollow shaft to impinge on the seal face [20]. This technique measured the effect of temperature on viscous shear and film thickness. The disadvantage of this approach was that the seal rotated around a stationary shaft, generating considerable centrifugal forces. Another interesting test rig was developed by Belforte et al for measuring radial contact pressure and contact area [21]. The technique involved a uniaxial load cell attached between a longitudinally split shaft.

The test apparatus discussed in this paper combines many the measuring functions into one apparatus, to allow a range of parameters to be assessed during testing. This data can then be readily used in the evaluation and design of tribotronic seals.

3. DEVELOPMENT OF TEST EQUIPMENT

3.1. Method

It was important to fully explore the need for a test apparatus that incorporated a range of measurement features. This approach meant it should be possible to collect data from various sensors simultaneously working with a range of seals. An iterative design method was used in the development of the test apparatus. The major challenge was to come up with a design that exceeded the minimum stipulations in BS 7780-4. The final horizontal shaft design discussed in this paper is a product of six design iterations, having started off as a vertical test bench. Below are some of the issues the design had to address.

- Motor size – A ECTOP TEFC 1.1 kW motor was specified based on calculations for maximum torque generated at start up for a dry running seal.
- Variable speed – An Optidrive E2 single phase motor controller is used to variably control the shaft speed.
- Pressurized oil bath – The oil-side of the seal can be pressurised to increase the pressure gradient at the seal interface. The bath can be drained to simulate partially flooded or dry running conditions.
- The removable seal carrier was designed to allow fitment of a range of nominal seal sizes ranging from 40mm to 170mm in diameter. (See figure 3b). The added advantage is that a variety of seal designs (sizes and forms) can be fitted and trialled without major modification to the whole test bench.
- Fittings and Instruments – the design allows for the fitting of sensors and actuators as the project progresses. An adjustable mounting on which to attach a CCD camera, to visualize the seal-quartz shaft interface, is incorporated.
- Torque measurement – A Magtrol TMB in-line torque transducer is incorporated in the drive train complete with flexible couplings to aid alignment.
- Leakage measurement – A leakage bin is mounted below the seal on the airside to measure the leakage rate.
- Eccentricity adjustment – The seal carrier can be positioned to cause a desired offset from the shaft rotational axis.

3.2. System design

Figure 3 and 3a show the general arrangement of the test apparatus. The system consists of a 1.1 kW 3-phase AC motor controlled by a 3-phase AC converter with a maximum rating of 2800 rpm. The motor is coupled to a Magtrol TM309 20 Nm torque sensor (accuracy <0.1%), complete with an integrated conditioning electronic module with a 0-10 VDC output.

The shaft is designed to run inside the seal which is held in a movable housing. The seal housing can be translated perpendicularly to the shaft to imitate axial shaft misalignment. A digital micrometre attached to the seal housing effects the offset. The shaft runs in an oil bath that can be pressurised and also incorporates heating elements. The induced pressure difference enables study of the effect of pressure difference, between the air and oil side, on oil film's load capacity. Facility to change the lubricant temperature assists with study of the effects of viscosity on the sealing function. Initial testing will incorporate a glass/acrylic shaft with a camera mounted on the interface. Steel shafts of varying surface finishes can also be mounted for testing in different conditions.

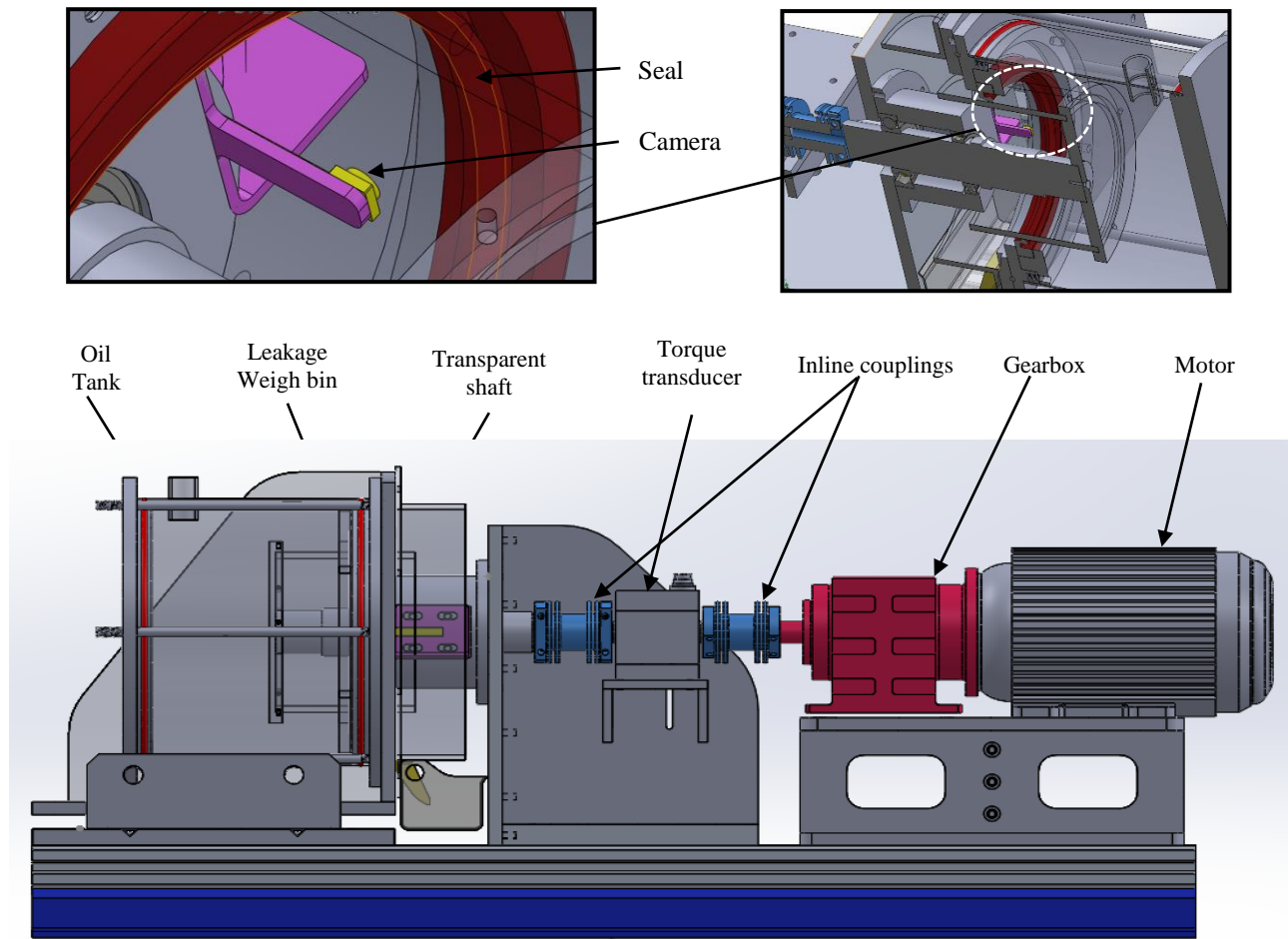


Figure 3 General arrangement of test apparatus. Oil tank containing the seal can translate perpendicular to the shaft axis.

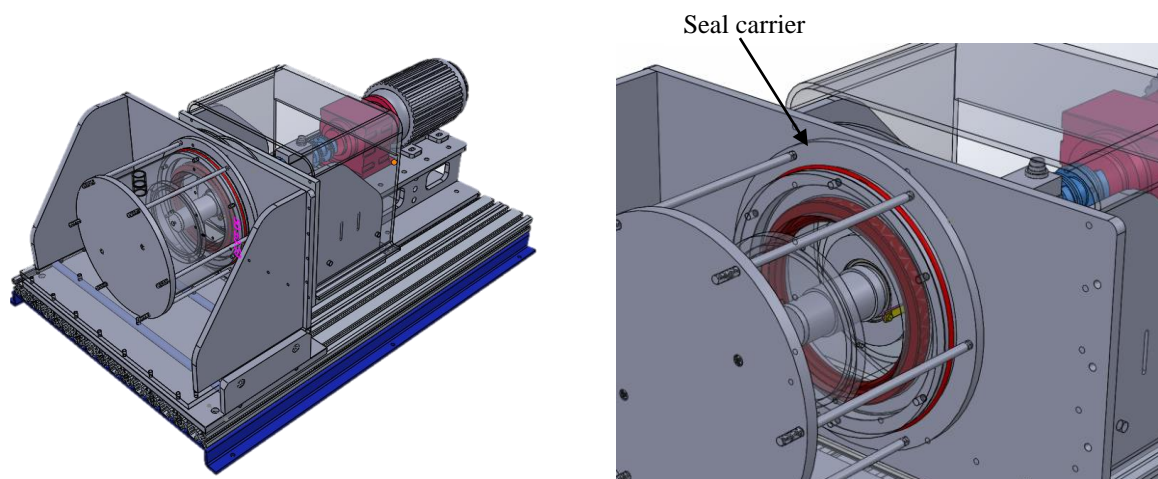


Figure 3. (a). Isometric view of test apparatus

(b). Removable seal carrier

The test apparatus is capable of parametric testing that is required to develop sensing and actuating the seal. Table 1 below shows some of the apparatus's capabilities. It also shows some of the governing equations and expected outputs.

Table 1: Features that are measurable with the new test rig.

| Feature | Governing Equation | General description | Measurement Domain | Output |
|----------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Hydrodynamic Pumping | Reynolds: $\Delta.(h^3/\mu)\Delta P = 6U \frac{dh}{dx}$ | Measurement of leakage (pumped fluid) against a given pressure over a specified period. | Time, lubricant volume, temperature, speed, shaft diameter. | Pumped volume as function of fluid viscosity (temperature and pressure normalised by diameter). |
| Eccentric Load | $[k]\{d\} = \{r\}$ | Adjustable Lip seal carrier capable of specified eccentricity. In-line rotary torque sensor. | Force, time, displacement, torque. | Eccentricity-load curve as function of temperature, speed, time. |
| Vibration | Time-Frequency domain analysis techniques. | Triaxial-Sensor-Acceleration due to inertial forces and mechanical excitation from seal. | RMS, acceleration, noise, displacement, time. | RMS curves as function of load, speed, pressure, displacement. |
| Seal-shaft interface | | Full HD 1080p Video camera | | Images-time, deformation at interface. |
| Oil Film thickness | Elastohydrodynamic lubrication theory $h_e = f(\eta u)^{0.66}$ | Calculated from lubricant viscosity in the contact, frictional torque and shaft speed. | Time, temperature, load and shaft speed. | Film thickness as function of speed, time, and temperature normalised for diameter. |
| Temperature | $\Delta T_{ss} = \frac{Q \rho C}{2\pi K r} = \frac{Q}{2\pi k r}$ $q_{total} = \mu p U$ | Calculation of flash temperature due to stationery heat source. Heat generated per unit area of contact in motion. | Time, temperature, load and shaft speed, pressure, contact area. | Changes in temperature at the contact interface. |
| Rotating Speed | $\omega_{max} = \frac{2\pi}{nK_p T_{clk}}$ | Speed measurement using Hall sensor. | Time, distance. | RPM, time, distance. |
| Pressure | | Positive pressure gauge. | Pressure. | Pressure. |

4. FURTHER WORK

This paper presented a general background on mechanical lip seals and the design of a novel seal test apparatus. The test apparatus is the first step in the development of an active seal which can autonomously monitor, change and communicate its condition in operation. Further publications will describe seal development and testing. The goal is to develop a seal that is self-powered, fully autonomous and uses the internet of things to improve condition monitoring and positively contribute to proactive maintenance.

Prior to any testing a full commissioning exercise will be carried out. This includes validation and verification of the test apparatus. The various sensors and measuring equipment relating to the test bench will be calibrated during this exercise. A battery of baseline tests will be conducted to gage the repeatability and reproducibility of the tests before actual testing commences.

The principal function of a seal is to prevent or control the flow of a fluid at a limiting boundary. In maintaining its sealing function faults may arise which lead to high friction (increased power loss) or wear (leading to leakage). It is possible to address both of these faults using tribotronics. The authors have chosen to address the problems of friction and wear by considering the shaft-seal interplay when it operates one of four states. These are:

State one: Leakage occurs at a stationary point on the seal and the shaft and seal axes coincide.

State two: Leakage occurs at a stationary point on the seal and there is an offset between the shaft axis and the seal axis.

State three: Leakage occurs at a moving point on the seal and the shaft axis and seal axes coincide.

State four: Leakage occurs at a moving point on the seal and the shaft axis and seal axis are not coincident.

Dividing the operating condition of the shaft and seal in this way allows incremental development of the various elements of the tribotronic system to be considered in a structured way against operating conditions which apply a developing level of demand. Selection of components for the tribotronic system is also being addressed in a structured fashion. System components will include: sensor(s), embedded intelligence in the form of software running on a small on-board computer, actuator(s), a communication system and a power supply.

Nomenclature

| | |
|-------------|--------------------------------------------------------------------------|
| C | specific heat capacity |
| h | fluid film thickness, separation of the mean lines of two rough surfaces |
| h_e | effective film thickness |
| K | thermal diffusivity |
| Kp | pre-scaler value for the Capture unit time base |
| k | thermal conductivity |
| n | number of pulses in one revolution |
| P | fluid film pressure |
| p | contact pressure |
| r | radius of asperity |
| Q | internal heat generated per unit volume |
| q_{total} | heat generated per unit area of contact |
| T_{ss} | absolute temperature at steady state |
| U | relative sliding velocities |
| x | contact width |
| μ | coefficient of friction between rubber and steel surface. |
| η | lubricant viscosity |
| ω | shaft speed in radian/sec |
| T_{CLK} | CPU clock period in sec |

REFERENCES

- [1] <https://www.technavio.com/report/mechanical-seals-market>: Accessed : 18/12/2016
- [2] Muller, H.K., Nau, B.S, (1998). Fluid Sealing Technology .New York: Marcel Dekker
- [3] https://www.google.co.uk/search?q=grindfos+chapter+5&oq=grindfos+chapter+5&aqs=chrome..69i57j0l3.17755j0j8&sourceid=chrome&ie=UTF-8#q=http://machining.grundfos.com/media/16611/shaftseal_chapter5.pdf Accessed : 18/12/2016
- [4] Glavatskih, S., Höglund, E. "Tribotronics – towards active tribology," Tribology International, vol. 41, N 9-10, 2008, pp. 934-939.
- [5] Sherrington, I., Glavatskih, S. "Application of TRIZ to the design of tribotronic systems," Proceedings of the 13th Nordic Symposium on Tribology, NORDTRIB 2008, 2008.
- [6] Salant, R. F. Theory of lubrication of elastomeric rotary shaft seals. Proc. IMechE, Part J: J. Engineering Tribology, 1999, 213, 189–200.
- [7] Johnston, D.E., Using the frictional torque of rotary shaft seals to estimate the film parameters and elastomer surface characteristics. In Proceedings of the Eighth BHRA International conference on Fluid Sealing, Durham.1978, paper C1 (BHRA, Cranfield, Bedfordshire).
- [8] Gabelli, A. and Poll, G. Formation of lubricant film in rotary sealing contacts: part 2 – a new measuring principle for lubricant film thickness. Trans. ASME J. Tribol., 1992, 114, 290–297.
- [9] Gabelli, A. Micro-elastohydrodynamic lubricant film formation in rotary lip seal contacts. In Proceedings of the 15th Leeds–Lyon Symposium on Tribology, 1989, pp.
- [10] Salant, R. F. and Flaherty, A. L. Elastohydrodynamic analysis of reverse pumping in rotary lip seals with micro asperities. Trans. ASME J. Tribol., 1995, 117, 53–59.
- [11] Poll, G. and Gabelli, A. Formation of lubricant film in rotary sealing contacts. Part II: a new measuring principle for lubricant film thickness. Trans. ASME, J Tribology.1992, pp 55-77.
- [12] https://www.google.co.uk/search?q=grindfos+chapter+5&oq=grindfos+chapter+5&aqs=chrome..69i57j0l3.17755j0j8&sourceid=chrome&ie=UTF-8#q=http://machining.grundfos.com/media/16611/shaftseal_chapter5.pdf: Accessed 20/03/2017.
- [13] Stakenborg, M.J.L., On the sealing mechanism of radial lip seals. In Tribology International, 21(6), 335-340.
- [14] Hidrovo, C.H., and Hart, D.P., Development of a dual purpose technique for the study of rotating shaft seals. In Proceedings of the 3Rd ASME/JSME Joint Fluids Engineering Conference, 1999, FEDSM99-7264.
- [15] http://www.freudenberg.com/FCODownloads/Freudenberg_Geschichtsbrosch%C3%BCre_en.pdf: Accessed : 18/12/2016
- [16] <https://www.google.com/patents/US20060006601?dq=freudenberg+seals+sensors&hl=en&sa=X&ved=0ahUKewjz3quxxKzUAhXrAsAKHQUUAU04ChDoAQgjMAA>: Accessed : 18/12/2016
- [17] <https://www.google.com/patents/US20080150238>: Accessed : 18/12/2016
- [18] 24. Simmering MSS1+Condition Monitoring, SIMRIT, [Product Brochure], http://www.koryosmt.co.kr/Front/service/download/2_simrit/2_Product%20Brochures/Simmerring%20MSS%201%20Plus.pdf: Accessed : 18/12/2016
- [19] Ogata, M., Mori, A. and Tagawa, N., Measurements Of EHL film creation in a lip type oil seal with axially ridged elastomer lip...(http://www.oetg.at/fileadmin/Dokumente/oetg/Proceedings/WTC_2001_files/html/M-15-06-530-OGATA.pdf). Accessed : 05/07/2016.
- [20] Stakenborg, M.J.L. (1998). On the sealing mechanism of radial lip seals. Tribology International,21(6).335-340.[DOI:10.1016/0301-679X(88)90110-7

An Intelligent Maintenance System for Driverless Vehicles

J. T. Philip, A. Onsy, M. R. Varley

Jost Institute for Tribotechnology, School of Engineering, University of Central Lancashire, Preston PR1 2HE, UK

ABSTRACT

The UK Department of Transport states that most road accidents are caused by human error. Several car manufacturing companies are now involved in research that aims to address this issue through the development of commercial driverless vehicles employing intelligent systems. Driverless vehicle research has been principally concerned with satisfying the requirements of level five of the Society of Automotive Engineering (SAE, USA) J3016 automotive automation standard; a standard that will improve commuter satisfaction and vehicle performance. However, many outstanding issues still need to be addressed, such as on-road safety and integration. Significant efforts are being made to advance current technology in order to create cost effective and robust driverless technology, which is expected to increase in the coming years. Additionally, an Intelligent Management System for Driverless Vehicles (IMSDV) is becoming a necessity as it is an intricate combination of Advanced Driver Assistive Systems (ADAS) management with an added intelligence algorithm that consists of various decision-making parameters, depending on a priority-based hierarchy. This resulted in research being conducted which focused on driverless vehicles in order to create a platform for testing, by developing a ‘Driverless Pod’ with intelligent systems suited for the advance of a complete IMSDV.

The paper introduces a new ‘Driverless Pod’ that has been developed and evaluated as an experimental test bench used to validate an IMSDV. The Pod incorporates several subsystems including different sensors, actuators and controllers, all of which are similar to those used in commercial driverless vehicles. Current development integrates both hardware and software. Further publications will detail two monitoring subsystems for the driverless vehicle steering system and vehicle wheel/tyre condition. A new and intelligent algorithm for driverless vehicle management that addresses health, safety and maintenance issues in relation to driverless vehicles will then be developed. These developments will be validated in three stages: laboratory testing, workshop testing and on the Driverless Pod.

Keywords: Multi-Disciplinary; Advanced Driver Assistance Systems (ADAS); Intelligent Management for Driverless Vehicle Systems (IMSDV).

Corresponding author: Jephin Thekemuriyil Philip (jphilip1@uclan.ac.uk)

1. INTRODUCTION

Since the conception of Intelligent Transportation Systems (ITS) in the 1980s, many transport researchers have progressed to work on the development of incident management models and integrated systems for real-time operations [1]. Several attempts have been made to develop a successive fully autonomous vehicle and many milestones have been reached. The challenges presented by the Defence Advanced Research Projects Agency (DARPA) in 2004 were perhaps the major spur behind the research for autonomous driving and autonomous vehicles, making it to be currently among the most intensively researched and publicly followed technologies in the transportation domain [2], which is an achievement. The challenges were introduced in order to emphasise the development of fully-autonomous ground vehicles. The agency cordially invited major companies and research organisations, and over 50 robotic and non-robotic vehicles drove simultaneously, but only six of these vehicles successfully completed the race. While the DARPA

Challenge remains the major demonstration of autonomous vehicle technology to date, it excluded many capabilities and requirements critical for actual city driving, thus, completing an off-road course within a time limit which remained a challenge [3].

Principal automotive companies (Ford, GM, Nissan, Volvo, etc.) and technology companies (Google, Induct etc.) have already demonstrated autonomous driving through working prototypes and pilots. The automotive industry is aware of the potential market emerging from autonomous driving. Relative companies are competing to prepare themselves for future revenue due to increased potential and other beneficial factors. Individual consumer benefits as well as societal benefits are a leading factor in the immensely gained attention of the research.

Current statistics concerning the number of accidents on the roads are worrying. Cars are a primary mode of transport and it is estimated that there are over 800 million cars on the road worldwide [4]. Reports produced by the UK Department for Transport state that 90 percent of UK road accidents are due to human error which include: speeding (32 percent of accidents), alcohol (21 percent) and distraction (17 percent). The average driver in England spends 235 hours driving every year, representing the equivalent of six working weeks. Additionally, according to the World Health Organisation, road traffic injuries account for almost 1.3 million deaths a year across the globe. Autonomous driving could help prevent these accidents by eliminating the role of human error in driving [5].

Several Advanced Driver Assistance Systems (ADAS) such as active lane keep assist, adaptive cruise control and self-parking are already currently available on the market have been combined as a major thriving aspect for the development of driverless vehicles. The replacement of a human by technology requires the performance of critical sensory functions using various technologies simultaneously. Many of these facilities currently exist in the latest technology, facilitating 'Level 3' SAE J3016 automotive automation standard, therefore requiring a high level of training in order to be viable for consumer use [6]. Several areas require mastery: vehicle location, prediction and decision algorithms, real-time accuracy and other factors; because technology must perform better than human eyes, ears, memory and coordination. This requires a high level of advancement. The UK has diagnosed the key areas which need consideration: safety and integration (Greenwich), vehicles on roads (Milton Keynes & Coventry), legal and insurance (Bristol).

The persistent transformation through the stages of autonomous driving is consistent with the latest technology, although mastery is not yet achieved in all areas. It is practically necessary to monitor the complete vehicle condition through an Intelligent Management System for Driverless Vehicles in order to contribute to health and safety research and the integration and vehicles on roads for driverless vehicles. The aim of this study is to find key areas which need substantial attention to health and safety.

To validate the IMSDV, a 'Single Seat Driverless Pod' has been developed. Current development integrates both hardware and software which are similar to those used in commercial driverless vehicles. Preliminary test results for the Driverless Pod have been included and indicate that the Pod is suitable for gathering real data that can be used in the final validation stage. Further publications will discuss the development of two monitoring subsystems addressing health and safety concerns for IMSDV.

2. DESIGN AND IMPLEMENTATION OF A NEW 'DRIVERLESS POD'

The aim of the research is to develop an Intelligent Management System for Driverless Vehicles. The research investigates new possibilities of contributing to the largely researched field of driverless vehicles, incorporating both design methods and computational systems that would allow the development of an Intelligent Management System for Driverless Vehicles (IMSDV) with the aid of a 'Driverless Pod' regarding current technological developments. The rationale behind the development of the 'Pod' is not to replicate the advances of a driverless vehicle, but to use as a test apparatus for gathering real data in order to develop an IMSDV.

A methodological proposal has been established in order to achieve the aim of the research. The design structure for the project research was considered prior to the commencement of the initial stages which include: study the background and existing systems, analyse current issues and areas which require further improvements, design and develop a driverless pod to validate the management system, and developed driverless pod will be used to test/ validate.

The proposed test apparatus (Driverless Pod) needed considerable attention to its structure as it involves combinations of multidisciplinary systems including, electrical, mechanical, control and other factors. Considering these subsystems which create the Pod, some specifications were considered:

- Dimensions: the structure of the Pod needs to be within the test bench architecture for ease of usage and for testing within a lab environment
- Pre-existing structure or build from first principles: rather than creating a structure/chassis from first principles, a pre-existing system that could be ambiguously fitted with driverless capabilities would be ideal
- Sensors: it should be possible to mount numerous sensors around the Pod; real-time data through sensory feedback and automated actuation
- Monitor various parameters such as movement through sensory fusion, actuators, and data processors

Many different possibilities were considered. After numerous considerations, concepts of building the mechanical structure from first principles were eliminated and the focus was towards electrically-powered karts because of their many advanced features which are suitable for the research. After assessing many options, a used mobility scooter was purchased and modified.

The Landlex Broadway mobility scooter was selected for the research of driverless vehicle systems. This remains an ideal platform, providing a robust chassis, multi-terrain tyres and user-friendly console with automatic breaking. Additionally, the ability to revert to human control of the vehicle if the software or the power fails is achieved by switching from automatic to manual mode. This pre-existing feature was greatly appreciated when selecting this scooter.

Additional modifications to the pre-existing structure were undertaken in order to fit the requirements of a 'Driverless Pod'. The Pod's steering was a major aspect to be considered during the redevelopment. Considering the weight of the scooter, no DC motor would operate efficiently. Various actuators were attempted and tested using numerous mounting strategies. Following the testing of various actuators, torque force was a recurring issue on each occasion. Therefore, replicating an actual car, an electric power steering system was fixed with the reasoning of high load and torque capabilities. A Vauxhall Corsa C, electric power steering (EPS) was modified to fit the Pod's chassis. Additionally, numerous mounting methods were used. The design of the mounting bracket and all subsequent parts have been manufactured to fit the Pod.

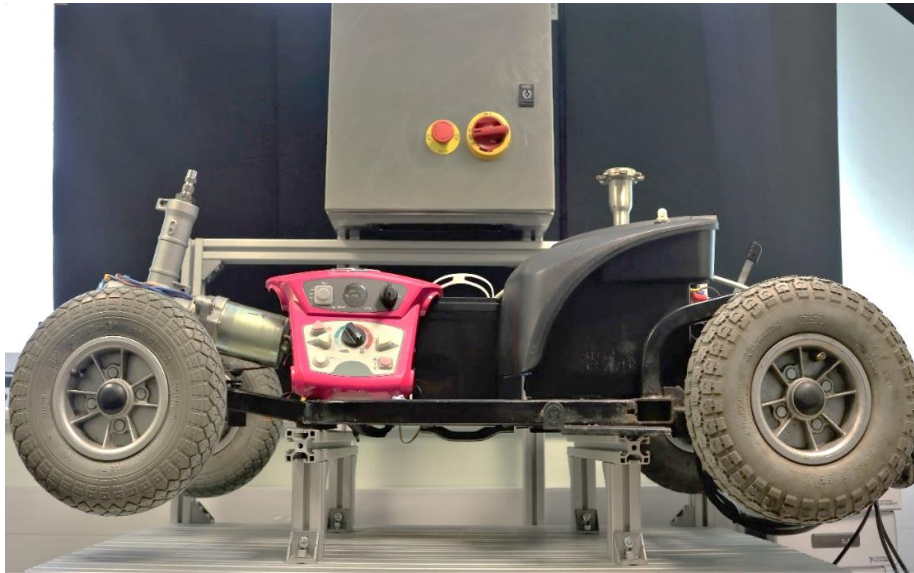
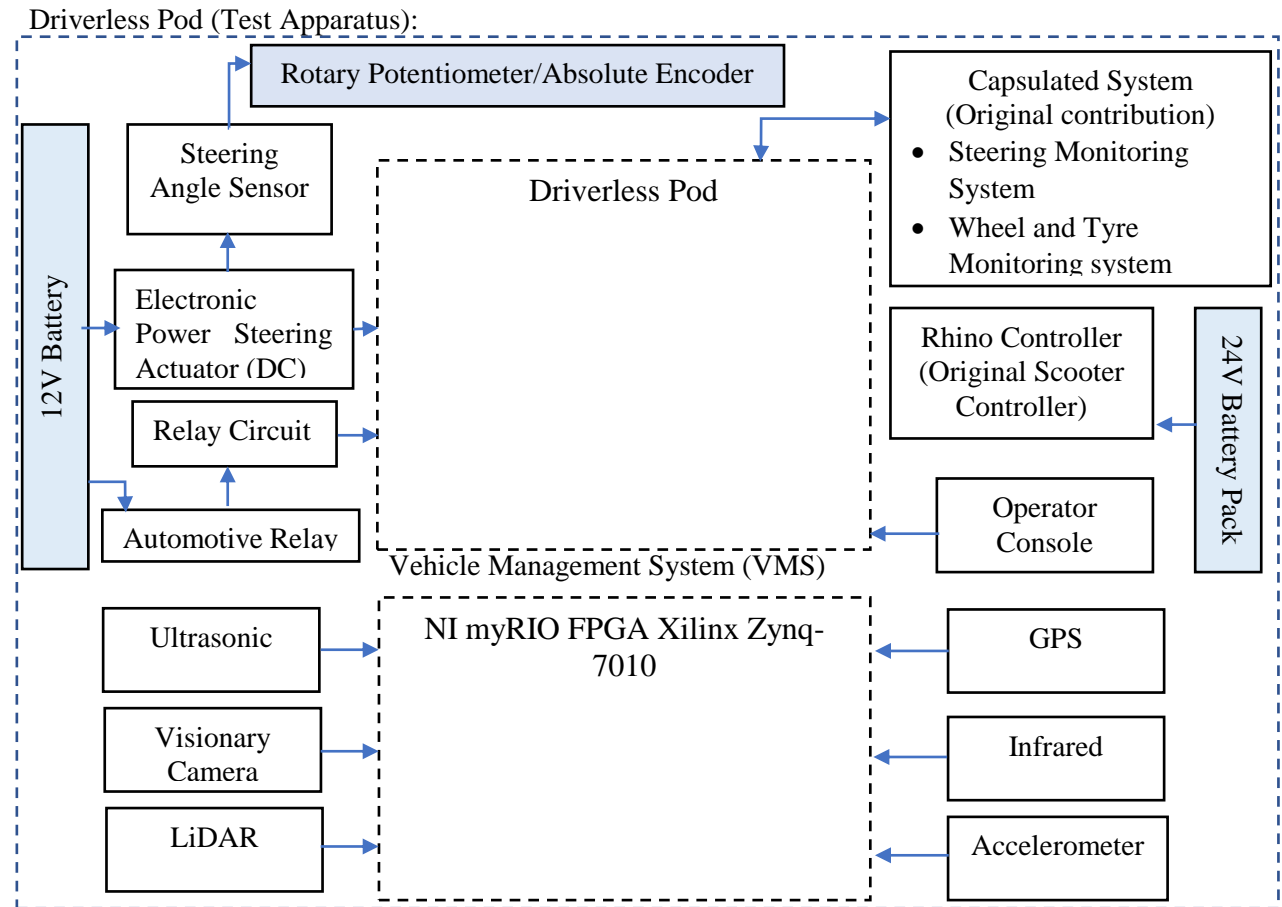


Figure 1. Landlex RS Broadway Mobility Scooter was purchased and modified to suit the requirements of a suitable test apparatus. The Pod incorporates multiple sensors, actuation and controller systems available in current driverless vehicles.

Patenting for ‘Autonomous vehicle arrangement and method for controlling an autonomous vehicle’ [7] states that it includes a receiving unit for travel order(s) and route planning with an array of sensors for detecting position, condition features, collision avoidance and a unit for controlling the vehicle actuator systems based on feedback generated by the vehicle control system. The array of sensors should include at least one range sensor at the front and rear of the vehicle, ultrasonic and/or microwave radar sensors arranged around the side of the vehicle, and at least one camera located in each of the front and rear areas of the vehicle.

The developed Pod was implemented with multi-sensors and actuators interfaced with an NI myRIO FPGA control module. The NI myRIO enables a completely powerful hardware tool which delivers the performance of a complex real-world system. NI myRIO features the Xilinx Zynq-7010 all-programmable system on a chip, which includes a dual core The ARM Cortex-A9 processor, and an Artix-7 FPGA controller processes the functionalities of the ‘Driverless Pod’. The current suite of sensors includes: video camera, infrared, ultrasonic, LiDAR, potentiometer, rotary encoder, EPS actuator, servo motors and relays. The myRIO module was programmed using a LabVIEW (Laboratory Virtual Instrument Engineering Workbench) interface to the controller to navigate the forward/ backward/left/right movements of the Pod remotely by using wireless communication.

In order to test the functionality, a pre-programmed sequence was created, implanting structured loops and case structures consisting of various moves and turns, each loop occupying specified time delays. Therefore, real-time sensory feedback was used to stop the pod from crashing and avoiding objects. Figure 2 illustrates an overview of the development of the ‘Driverless Pod’ which includes the multiple sensors, actuators and controller diagram.



3. TESTING AND EVALUATION

To evaluate the developed ‘Driverless Pod’s’ performance under different conditions and to verify the analytical results, it was subjected to test procedures. These include a two-stage procedure which tests and validates the operation of each subsystem individually and also the developed Pod to confirm its suitability for use in constructing an IMSDV.

3.1. Evaluation of sensory actuation system

The developed sensory actuation system consists of two main actuators, the main driving motor and the steering motor. Each motor has been assigned with subsequent sensors which acquire different parameters relating to distance, speed and angle. Therefore, the implemented sensors and actuators have been individually tested and validated. Initial results have been acquired using the LabVIEW platform and subsequently documented.

3.1.1 Infrared Range Finder

The Sharp 2Y0A21 IR Ranger Finder was used to identify the range between the obstacles and the Pod. To calibrate the sensor accurately, multiple measurements were logged and compared. By mounting the sensor to a vertical surface, with a known range, an object was placed in front of the sensor and moved respectively to a known range on each occasion in order to identify the resulting output, V_o .

The measurable range for the sensor was 10 to 80 cm, therefore, it could be noted that at the specified range, a typical analogue voltage output was 2.30 to 0.4 V with respect to 38.3 ± 9.6 ms measurement time, 5 ms update lag.

$$R = K_s \frac{1}{V_o} + K_o$$

R = Range, K_s = Calibration Coefficient Scale, V_o = Voltage Output, K_o = Calibration Offset

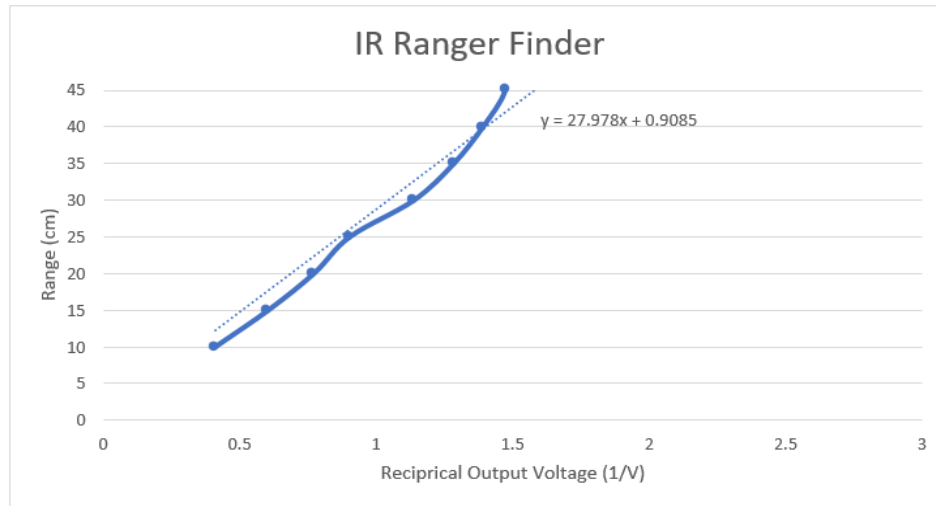


Figure 3. The graph shows the voltage in reciprocal form versus range (cm). A linear region is identified and extracted with a linear trend-line after calibration.

It can be concluded from this test that the scale factor $K_s = 27.978$ cm-V and offset, $K_o = 0.91$ cm. At close ranges, the sensor has improved sensitivity as opposed to objects further away. Also, the sensor must be at least 10 cm from the target for accurate measurements within the range.

3.1.2 Main Driving Motor

The motor was controlled wirelessly and the NI myRIO FPGA module could acquire, display and log data. The logged data could be plotted in order to analyse the trend-line. This would be utilised while testing and validating the IMSDV, using the Pod (test apparatus) through wireless connectivity. The Pod was wirelessly controlled using the FPGA controller, interfaced with the original mobility scooter controller, in order to test and evaluate forward/reverse movements using a developed code through LabVIEW. The actuator was deployed by using digital I/O with additional relay circuitry.

3.1.3 Steering Angle Sensor

The Pod's steering is achieved by using Electric Power Steering (EPS) together with an angle sensor. A rotary potentiometer has been used for preliminary testing and the relationship between the acquired voltage and the position angles was noted. Therefore, the sensitivity equation was generated in order to identify the variation between central position/left/right angles.

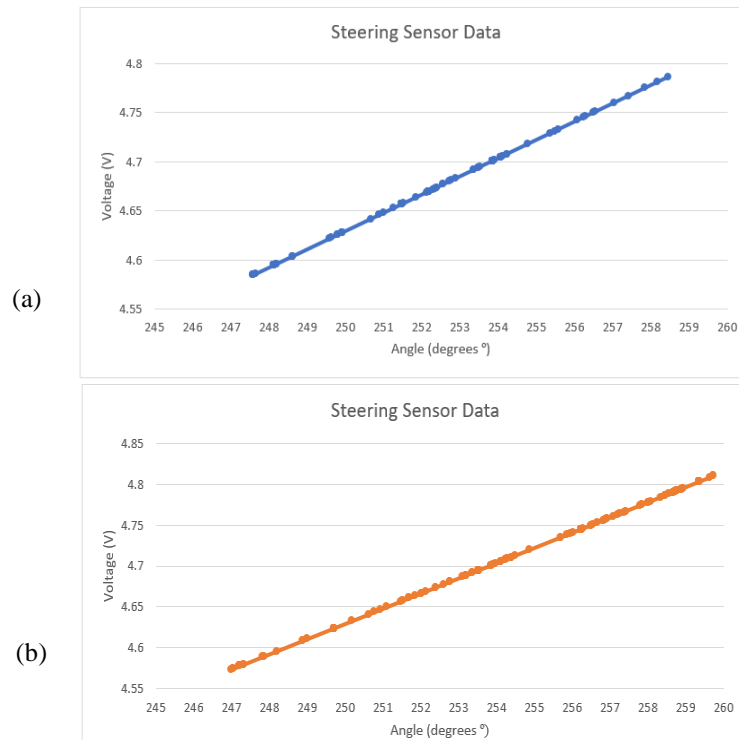


Figure 4. Graphs (a – left turn, b – right turn) illustrate the relation between Voltage Output and Steering Sensor angle from a test run of the Pod. This relation was acquired using the above equations. The test run of the Pod consisted of controlling the steering system wirelessly through LabVIEW FPGA controller.

3.2 Wireless Control

Wireless connectivity was achieved using the NI myRIO FPGA module via a laptop. This allowed remote access to the Pod via the subsequent circuitry. A relay board interfaced with the controller with an additional battery source was used for the forward/backward/left/right movements. Additional sensors have been interfaced directly to the FPGA.

Following validation of the developed Pod's installation and the controlling of the Pod remotely, a pre-programmed sequence was created, positioning structured loops and case structures consisting of various moves and turns creating a sequence, each loop occupying specified time delays. Therefore, real-time sensory feedback was used to prevent the Pod from crashing and to avoid objects.

4. DISCUSSION

The research investigates new possibilities of contributing to the largely researched field of driverless vehicles; incorporating both design methods and computational systems that would permit the development of an Intelligent Management System for Driverless Vehicles. The paper includes the development of a 'Single Seat Driverless Pod' to be used in the actual system validation. Succeeding work will include new monitoring systems with a further focus on health and safety; this is an area which needs to be addressed.

According to SAE the steering system in an autonomous driverless vehicle is one of the current development tracks [6]. Vehicle steering is considered to be at level 2 of the SAE J3016 automotive automation standard, so that it incorporates human interaction as part of the decision. Current fully automated power assisted or semi-assisted steering systems utilise feedback signals by using angular position sensors to enhance the driving experience.

Defective wheel bearings are an additional cause of an uncomfortable drive. The monitoring of the life of wheel bearings could be used to develop advanced maintenance strategies for driverless vehicles using a wireless monitoring system. Wheel bearing defects will create excessive vibration and such signals can be monitored by using piezo electric elements, and by providing warning systems. This is built on the study of 'A New Acoustic Emission Wireless Monitoring System; An Experimental Validation of Bearing Condition Monitoring' [8].

Tyre pressure has a significant effect on vehicle performance and therefore, tyre life, correct tyre pressure and temperature should always be revised. The life of the tyre can be extended by monitoring tyres with pressure sensors, thereby ensuring that the pressure is always kept at the standard limits.

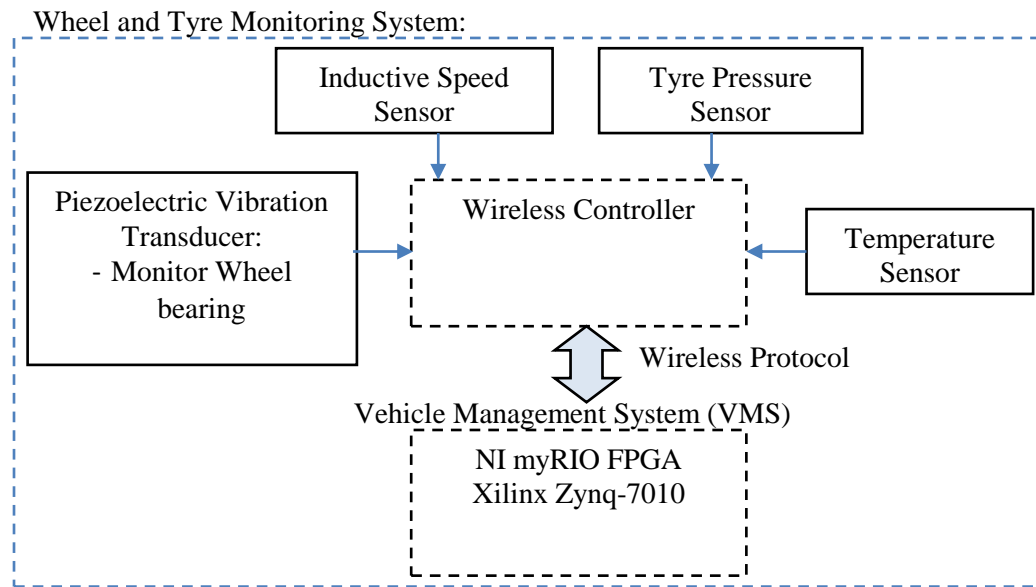


Figure 5. Sub System Block Diagram – Wheel and Tyre Monitoring System.

5. CONCLUSIONS

Driverless vehicle research has been principally concerned with satisfying the requirements of level 5 of the Society of Automotive Engineering (SAE, USA) J3016 automotive automation standard; a standard that will improve commuter satisfaction and vehicle performance. However, several outstanding issues still need to be addressed, such as on-road safety and integration. Therefore, this study addresses the ongoing development of an Intelligent Management System for Driverless Vehicles.

This paper documents the work achieved in the initial developments of the IMSDV. It presents critical reviews of influential literature on driverless technology that is associated with this field, in order to highlight the existing theoretical and technical gaps within the international research. Combination of both design and development aspects have been considered in this paper in order to develop an adequate test apparatus (Driverless Pod) that suits the requirements.

The authors have presented preliminary test results for the Driverless Pod including the use of multiple sensory actuation systems which indicate that the Pod is suitable for gathering data that can be utilised in the final validation stage of IMSDV. These developments will be validated in three stages: laboratory testing, workshop testing and on the Driverless Pod. The following publications will illustrate the integration of monitoring systems for driverless vehicles by utilising IMSDV.

REFERENCES

- [1] Ozbay, K. and Kachroo, P., 1999. Incident management in intelligent transportation systems.
- [2] L. C. (Ed.). (2016, May). Transformational Technologies in Transportation (Rep. No. ISSN 0097-8515). doi: <http://onlinepubs.trb.org/onlinepubs/circulars/ec208.pdf>.
- [3] Buehler, M., Iagnemma, K. and SINGH, S., 2009. The DARPA urban challenge: autonomous vehicles in city traffic. Springer.
- [4] Thrun, S., 2010. Toward robotic cars. *Communications of the ACM*, 53(4), pp. 99-106.
- [5] Department of Transport, 2015-last update, The Pathway to Driverless Cars Summary report and action plan [Homepage of Department for Transport, London: ISBN 978-1-84864-153-2], [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/401562/pathway-driverless-cars-summary.pdf.
- [6] SAE, T., 2013. Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems-J3016. Society of Automotive Engineers: On-Road Automated Vehicle Standards Committee.
- [7] Bergholz, R., Timm, K. and Weisser, H., 2000. Autonomous vehicle arrangement and method for controlling an autonomous vehicle.
- [8] Onsy, A., 2013. A New Acoustic Emission Wireless Monitoring System; An Experimental Validation of Bearing Condition Monitoring. *SAE International Journal of Aerospace*, 6(2013-01-2221), pp. 534-539.
- [9] One Box: Driver and Vehicle Data Management System Criteria. September 2012. 28/12(ISBN: 978-1-78246-002-2).
- [10] Department of Transport, September 2016-last update, Reported Road Casualties in Great Britain: quarterly provisional estimates Q2 2015 [Homepage of Department for Transport, London], [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/473850/quarterly-estimates-april-to-june-2015.pdf.
- [11] Fagnant, D.J. and Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 77, pp. 167-181.
- [12] Soltani, A., Goodarzi, A., Shojaiefard, M.H. and Khajepour, A., 2017. Developing an active variable-wheelbase system for enhancing the vehicle dynamics. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, pp. 0954407016681683.
- [13] Ali, E., Onsy, A., Elsherif, I. & Metwaly, M. (2015) Design and Development of a Low Cost RHex Robot. *Proceedings of the 2015 Conference of the Society for Machinery Failure Prevention Technology*, USA, Huntsville.
- [14] Sharp (2006) 02Y0A21 IR Ranger Finder Datasheet
- [15] Corsa owner's manual (2005) Electric Power Steering (EPS)

Session 15

Advanced Signal Processing and Big Data 2

Session Chair Dr. Mohammed Khan

The Way Cyber Physical Systems Will Revolutionise Maintenance

(E. Jantunen, U. Gorostegui, U. Zurutuza, F. Larrinaga, M. Alnano, G. Di Orio, P. Malo C. Hegedus)

Nuisance Attribute Projection Based Channel Compensation Technique and its Application in Bearing Performance Degradation Assessment

(H. Jiang, J. Chen, G. Dong)

Dynamic State Recognition Using CNN-RNN Processing Pipeline

(E. R. Anas, B. J. Matuszewski)

Study on Feature Extraction of Bearing Incipient Weak Fault Based on Dictionary Learning

(G. Dong, J. Chen, H. Yuan)

The Way Cyber Physical Systems Will Revolutionise Maintenance

E. Jantunen¹, U. Gorostegi², U. Zurutuza², F. Larrinaga², M. Albano³, G. Di Orio⁴, P. Malo⁴, C. Hegedűs⁵

¹VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Finland

²Mondragon University, Loramendi 4, 20500 Mondragon, Spain

³CISTER, ISEP/INESC-TEC, Polytechnic Institute of Porto, Portugal

⁴Dep. De Eng. Eletrotécnica, FCT-UNL, UNINOVA-CTS, Lisboa, Portugal

⁵AITIA International Inc., Czetz János str. 48-50., Budapest HU-1039, Hungary

ABSTRACT

The way maintenance is carried out is altering rapidly. The introduction of Cyber Physical Systems (CPS) and cloud technologies are providing new technological possibilities that change dramatically the way it is possible to follow production machinery and the necessity to carry out maintenance. In the near future, the number of machines that can be followed from remoteness will explode. At the same time, it will be conceivable to carry out local diagnosis and prognosis that support the adaptation of Condition Based Maintenance (CBM) i.e. financial optimisation can drive the decision whether a machine needs maintenance or not. Further to this, the cloud technology allows to accumulate relevant data from numerous sources that can be used for further improvement of the maintenance practices. The paper goes through the new technologies that have been mentioned above and how they can be benefitted from in practise.

Keywords: Condition Based Maintenance (CBM), Cyber Physical Systems (CPS), Overall Equipment Effectiveness (OEE), e-maintenance, Internet of Things (IoT)

Corresponding author. E. Jantunen (email: erkki.jantunen@vtt.fi)

1. INTRODUCTION

We are on the verge of a great deal of changes on how machine maintenance is performed, since the era of CPS has come. In fact, embedded devices are now part of every environment, comprising where industrial machinery is located. The online collection of data is able to empower big data algorithms to perform analysis and prediction on many aspects of life, comprising the condition of an industrial machine in the present and in the close future. This can allow for managing maintenance in novel ways. E.g. depending on the business case, it can be reasonable to predict the remaining lifetime of a machine, and buy the correct spare parts in advance to minimize the downtime of the machine. The application of CPS is a consequence of the direction where industrial automation is going. Currently, it is common to organize operations related to industrial automation into a pyramid, e.g. as per ISA-95 standard [1]. Even though there is some difference between ISA-95 compatible formalizations, a coarse definition of its levels can be given: Level 0 (field level), composed of sensor and actuators that interact directly with the process or the machine. Level 1 (direct control) is related to the operations performed by special-purpose hardware, such as PLCs, industrial pcs, DSP processors, as drivers of the actuators and collectors of sensors data. Level 2 (Supervisory Control) acts online with respect to the industrial process, and its major functions are to allow an operator to change set points for the industrial process, and monitor the activity of the machine(s) whose data is collected. Level 3 (Production Control) is targeted to the manufacturing operation as a whole, and includes maintenance, production, quality assurance, inventory management. Level 4 (Enterprise Control) consists mainly of management functions, and it is used to drive the objective of the manufacturing process by scheduling its operations. CPS are located on Level 0, and receive support in the factory up to Level 3, where the process becomes distributed over internet. It can be considered that Production Control involves transporting data over larger distances, and a modern approach is apply the concept of “servitization”, where the upper levels are mediated and implemented as services, to maximize scalability and flexibility of the systems. In this context, the two main goals of data communication and processing can be supported by means of a communication middleware the first, and of fog computing the latter. Fog computing is an

evolution of cloud computing, where micro-clouds and edge servers are disseminated closer to where the data are collected and used, to allow for data preprocessing and local computation, to parallel the bulk computation performed in the cloud. The acceptance of these new approaches is hindered by a few issues, some of them on the technological side, and some of them on closer to the business. This paper aims to provide an overview of the novel maintenance approaches (section 2), of how embedded systems can support it (section 3), and of the servitization (section 4) of the maintenance-related operations (section 3). Later on, the paper analyzes the issues that must be overcome to reach the technology maturity level needed for the implementation of this technological revolution (section 5), and the business impact that it can provide (section 6). Some conclusions are drawn in section 7.

2. CONDITION BASED MAINTENANCE

CBM is a predictive maintenance strategy that is based on the continuous monitoring of various parameters of an asset to evaluate its health level and future development. The increasing popularity of predictive maintenance strategies is preceded with the necessity of an improved Overall Equipment Efficiency (OEE). The OEE is a term expressed in percentage that evaluates how effectively a manufacturing operation is utilized [2]. The metrics that are taken into account to calculate OEE are *availability*, *performance* and *quality*. The OEE is usually used as to identify the scope for process performance improvement [3]. The CBM approach enables an increase in the availability of the asset, thus, increasing the OEE by reducing the unplanned downtimes of the machines and saving costs in unnecessary repairs. The information gathered by the condition monitoring architecture allows the planning and scheduling of maintenance, while also reducing the storage needs for spare parts. This information can also help to identify failures and determine how to avoid them before occurrence. To be able to implement a CBM system, an initial analysis needs to be done in order to decide which parameters are worth monitoring. Different analytical tools such as Fault Tree Analysis (FTA), Root Cause Analysis (RCA) or Failure Mode and Effects Analysis (FMEA) are useful when identifying the root cause of a failure and the best way to avoid them. These tools allow to decide which parameters to monitor and consequently identify the main failure causes, and to improve future designs. The analysis also considers that its result can be a suggestion not to apply condition monitoring to said equipment. Moreover, there are some cases where the traditional corrective maintenance strategies are also valid, e.g. when a machine can still work when a component breaks down and can wait until the maintenance is scheduled to change the broken component. The Machinery Information Management Open System Alliance (MIMOSA) Open System Architecture for Condition Based Maintenance (OSA-CBM) is a not-for-profit association that develops the open information standards for Operation and Maintenance in manufacturing [4]. The standards that MIMOSA develop are compliant with the ISO-13374 Standard (Condition Monitoring and Diagnostic of Machines), being an implementation of the latter's functional specifications. According to this standard, a CBM system should be composed of various functional blocks: 1) Data Acquisition (DA), 2) Data Manipulation (DM), 3) State Detection (SD), 4) Health Assessment (HA), 5) Prognostics Assessment (PA) 6) Advisory Generator (AG)[5]. One of the main objectives of the MIMOSA OSA-CBM is to standardize the information flow between the various blocks, so that equipment from different vendors could be interoperable. The Data Acquisition block is responsible for picking up the physical phenomenon and converting it to a readable digital signal, by means of a transducer or a sensor. After applying some filters to reduce the noise level and amplify the signal if needed, the analogue signal needs to be converted to a digital signal through an Analog to Digital Converter (ADC) so that a computer or a processing system can manipulate the data and get meaningful information. The data that comes from the ADC is usually referred to as "raw data". This raw data is then sent to the Data Manipulation block. Here, mathematical algorithms such as Fast Fourier Transform, kurtosis or envelope analysis are applied. The outcome of this analysis is stored in the database and, depending on the application, even the raw data could be stored. In the State Detection step, data from the first and second blocks are compared with expected values, to verify if they fit into previously defined limits, generating alerts when they do not [6] [5]. The main objective of a CBM system is to make a diagnostic assessment on the health level of the asset and then do a prediction on how the state is going to degrade. This analysis is done in the Health Assessment and Prognostic Assessment blocks. Usually, the diagnostic is based on the health history trends, operational status and load history while taking into account possible faults. The prediction estimates the Remaining Useful Life (RUL) of the asset and it can be computed with three main approaches: model-based, data-driven and hybrid approach [7]. The model-

based approach tries to mathematically describe the physical phenomena of degradation. This method can be very accurate but it gets more complex while being more detailed. The data-driven model is based on a statistical approach, where pattern recognition and machine learning algorithms are implemented. One drawback of this method is that it may be quite time-consuming to get an initial data batch (training data), because usually run-to-fail data is needed. The hybrid approach takes the advantages of both methods and tries to create a model where physical phenomena are described but also uses statistical data to complement it and create more reliable and accurate outcomes. The last step of the chain -Advisory Generation- is to create a report stating the recommended maintenance actions that have to be taken in order to optimize the useful life of the product. These actions could be, e.g., to schedule a replacement of a component at the best possible moment, reduce the speed of a rotating machine to increase its life expectancy, or do nothing if the component is in a healthy state. Other references, such as [8] also state that there could be an extra layer or block, the Presentation Layer, where the information from all the previous step can be accessed from. However, usually only the information of the last three steps is displayed because it ends up being the most useful.

3. INTERNET OF THINGS AND CYBER-PHYSICAL SYSTEM OF SYSTEMS

CPS pertain to a new discipline – in the field of computer science and communication technologies – that is revolutionizing the way complex systems are designed, implemented and deployed. CPS are the “integrations of computational and physical processes” [9], in the sense they can integrate and merge the physical world and the virtual world. The term CPS was coined in 2006 by Helen Gill at the National Science Foundation in the United States and it naturally emphasizes the necessary link between physical and virtual that is frequently ignored in a world constituted by applications that run only on PCs. As stated in [10], today around 98% of microprocessors are embedded, and hence directly connected to the physical world by various means of sensors and actuators (sensing and acting on the environment) and increasingly connected with one another and the internet. This trend is expected to continue as also confirmed in [11], where the importance of networked distributed systems of embedded computers during the coming years is highlighted. However, as stated in [12], CPS means more than networking and information technology, this concept presupposes that information from the physical world (assets) needs to be integrated and used within the cyber world, while creating sustainable feedback loops, where decisions computed within the cyber world affect the physical world and vice versa. This is the main assumption for CPS and implicitly designates a multidisciplinary knowledge to establish the tight integration of perception, communication, learning, behaviour generation, reasoning for creating CPS-populated systems. As a matter of fact, CPS include a wide variety of systems in the most disparate contexts of application (e.g. aerospace, automotive, manufacturing, white goods, healthcare, telecommunications, power grid, etc.) and domains/fields of engineering (e.g. chemical, electrical, power, mechanical). Besides the domain and the application context, knowledge of the software, communication, control theories, methods, methodologies and tools is also required [13]. By synthesizing all the precedent assertions, the following core elements and/or characteristics can be identified for CPS, extended from [12], [14]: 1) Enhancement of physical entities with cyberspace capabilities; 2) Networked at multiple and extreme scale; 3) Dynamic behaviour (plug and unplug during operation); 4) High degrees of automation, leading typically to closed control loops; 5) High degree of autonomy and collaboration to achieve a higher goal; and 6) Tight integration between devices, processes, machines, humans and other software applications. The above core elements and/or characteristics of CPS spontaneously point to Internet-of-Things (IoT) technologies and System-of-Systems (SoS) approaches, methodologies and/or research initiatives as the backbone for creating Cyber Physical Systems-of-Systems (CPSoS), i.e. large, complex, often spatially distributed CPS that exhibit the features of Systems-of-Systems [15]. IoT solutions can provide the backbone infrastructure to enable seamless integration between the physical and virtual worlds, i.e. to enable the easy and rapid access to the physical world contents and events (e.g. information) by means of computers and networked devices according to the paradigm “anytime, anywhere” [16]. Conceptually, both IoT and CPS are networked systems needing high-degrees of automation that are likely to involve physical sensing and/or embedded devices i.e. both combine aspects of the physical and digital/cyber worlds. However, there is a slight difference between them. In fact, one usually refers to CPS in the case of systems/problems that involve large-scale real-time control (e.g. time critical problems); or problems that involve integrated control of combined organisational and physical assets by profoundly relying on their virtual representation (i.e. the modelled behaviour of

cyber entities; conversely). Meanwhile, IoT accounts for situations that collect and process sensor data (including IoT analytics problems) without essentially involving real-time control. CPS here would relate to systems involving collaborative automation of networked embedded systems, and tight human machine interactions; whereas IoT would relate to target systems/applications involving fewer collaborative automation and requiring internet connectivity only. The concept of SoS covers an entire research background constituted by relevant theories, tools, knowledge and approaches to analyse, design, model and control large distributed systems that consist of networked interacting elements [15]. Therefore, CPSoS are essentially ecosystems of CPS and IoT solutions that rely on the effective and efficient collection, provisioning, analysis and visualization of large quantity of data to monitor, diagnose, adapt and optimize (through reconfiguration) the overall behaviour in the environment where they are operating. Therefore, the availability of this large quantity of data immediately triggers the implementation of advanced monitoring strategies for assets management while facilitating – at the same time – the adoption of policies and strategies for maintenance, such as: 1) CBM); and 2) Proactive Maintenance (PrM). As a matter of fact (as deeply explained in [17]), the successful implementation of CBM and PrM strategies thereafter, will be possible as a result of the presence of an efficient and effective monitoring infrastructure that can gather relevant operational data from assets, combine and analyse these data to identify possible breakdowns and their root causes. Consequently, CPSoS have a tremendous potential in promoting the meshing of virtual and physical worlds while leading the interconnection of people, processes and infrastructures within interactive and responsive networks of CPS for fast evaluation of asset performances.

4. SERVICITIZATION BASED ON CPS: THE ROLE OF THE CLOUD AND FOG/EDGE COMPUTING

The wider dissemination of CPS and their aggregation into CPSoS is having disruptive effect on the market structures of enterprises, due to the highly-networked characteristic of most of the systems of today. CPSoS will change existing business models while enabling new suppliers of services for CPS-based systems to enter the market [10]. At the same time, the emergence of cloud computing (and more recently of fog/edge computing) is creating new and exciting opportunities for CPSoS by enabling: 1) The wider consumption of the data generated within the CPSoS ecosystems and the services provided [18]; and 2) The creation of new services as composition of the ones exposed within the CPSoS. As stated in [19], the rise of cloud computing [20] has initially created the foundations for breeding CPSoS by providing: 1) an infrastructure for CPS integration, i.e. services and/or atomic functionalities provided by CPS that are part of the CPSoS ecosystem can potentially be accessed/used over the internet by other CPS or applications; and 2) a huge amount of computational and storage resources that are available within the cloud and can be used “on-demand”. Service provisioning both at CPS and at CPSoS level, make maintenance more flexible, enabling remote monitoring and control of processes and tasks. Service management becomes key in this extended context and a number of duties must be addressed: 1) Monitor and control routing of message exchange between services, 2) Resolve contention between communicating service components; 3) Control deployment and versioning of services; 4) Marshal use of redundant services; 5) Cater for commonly needed commodity services like event handling and event choreography, data transformation and mapping, message and event queuing and sequencing, security or exception handling, protocol conversion and enforcing proper quality of communication services. An Enterprise Service Bus (ESB) is a software architecture model used for designing and implementing the interaction and communication between mutually interacting software applications and components in a Service Oriented Architecture (SOA) like the one emerging for CPS in the maintenance context. ESB motivation comes from the need to find a standard, structured and general-purpose concept for describing implementation of loosely coupled services that are expected to be independently deployed, running, heterogeneous and disparate within a network. The main functional areas for an ESB are [21]: 1) Architecture. The main issues covered in this area are support for fault tolerance, scalability and throughput, the ability to federate with other ESBs, the supported topologies, and features supporting extensibility. 2) Connection. The key features in this group include support for a wide range of messaging standards, communications protocols, and connectivity alternatives. 3) Mediation. This group deals with key requirements related to dynamic provisioning of resources, transformation and mapping support, transaction management, policy metamodel features, registry support, and service-level agreement coordination. 4) Orchestration. This layer provides lightweight orchestration of services and more-robust Business Process Execution Language (BPEL)

and/or Business Process Modelling Notation (BPMN) support. 5) Change and control. The main components in this group are design tooling, lifecycle management, technical monitoring, and security. The Arrowhead project [22] addresses efficiency and flexibility at the global scale by means of collaborative automation. Arrowhead assumes that a service-based approach will be the feasible technology that enables collaborative automation in an open-network environment connecting many embedded devices and CPS. The multi-billion device/service perspective places a very strong demand on the interoperability and integrity of devices and services provided by the multitude of players in the market place. Thus, Arrowhead's grand challenges are: 1) Enabling the interoperability of services provided by almost any device. 2) Enabling the integrity of services provided by almost any device. Next to cloud computing, fog/edge computing architectural pattern has been recently introduced. This pattern is aimed to extend the cloud computing paradigm to the "edge" of the network for those applications that are latency-sensitive and – thus – have strict delay requirements [23]. Therefore, fog/edge computing is about pushing intelligence, data analytics and knowledge generation into smaller clouds, near to physical devices. Traditional cloud computing data centres are also utilized, however, but moved closer to source of the data [24] while supporting localization, i.e. location awareness and distribution. Actually, fog/edge computing paradigm reflects better the complexity, heterogeneity and distribution of CPS and IoT populated systems and their ecosystems than cloud computing. The combination of cloud and fog/edge computing paradigms and CPSoS is allowing companies to evolve their hierarchical and static configuration into a new one, characterized by agility, openness and peer-to-peer-based interactions. This trend is also confirmed by the manufacturing industry – considered here as the reference sector to show the economic health and welfare of a country – where the combination of ICT and CPSoS is triggering the transformation of the manufacturing value chain patterns from pure manufacturing and selling physical products to the provisioning of sophisticated integrated solutions where physical products are enhanced by functions and services [25]. As explained in [26], this business trend can be designed as "servitization" that means the process of creating value in products and goods by adding services. The term was initially coined by Vandermerewe & Rada [27], and now is widely recognized and adopted to identify a specific competitive manufacturing strategy as pointed in [28]. During the production stage cloud-based CPSoS are facilitating the integration of the data within the enterprise, i.e. from industrial assets at the shop floor level to business applications at management level. In the pre-production and post-condition stages, cloud-based CPSoS can provide relevant data that can be used to support both the Product Lifecycle Management (PLM) and the Service Lifecycle Management (SLM). In this landscape, maintenance assumes a vital role in guaranteeing the perfect working conditions of the industrial assets, and thus the quality of the final product and a resilient service. However, the effective and efficient implementation of manufacturing strategies and procedures strictly depends of the availability of transparent and as much as possible exhaustive insights about industrial assets and products [29]. The creation of cloud-based CPSoS can lead the implementation of maintenance specialized platforms and frameworks to improve industrial assets productivity by relying on the right information at the right time.

5. ISSUES TO TACKLE FOR IoT- AND CPSoS-BASED CBM

In this paradigm, data from its sources travel far and might be processed several times before its information is properly extracted and utilized. This can be facilitated by a wide variety of communication protocols, technologies and architectures. How the (mostly sensory) data is then transmitted, virtualized (as per the CPS approach) and used raises higher level issues that might not have been considered previously for condition monitoring scenarios: 1) How to transmit these data from the physical system and to where? 2) How to create interoperable data representation and semantics? 3) What can be the backend that processes this inbound data streams in a scalable manner? 4) How can we still maintain real-time restrictions and abide by communicational constraints? Firstly, the acquired data that might be available with great time and value resolution. However, it is often not practical to be transmitted "as is" from the device or machine for communicational constraints. Therefore, low level pre-processing and local storage might be required. It can include sensor data fusion and filtering, elimination of noise and erroneous data and even could utilize advanced logics (e.g. rule-based notifications). The purpose of this is to only transmit an "extract" of the readouts that still represent the physical process sufficiently for later, higher level processing (within the cloud). It is also of essence that the cloud processing units should still be able to request the raw data streams on demand (if the RUL or RCA algorithms require them), besides the regular

bandwidth-friendly messages. This should be facilitated by the sensor nodes, gateways and the employed communication protocol itself. To this end, the traditional client-server architecture is being replaced with bi-directional, persistent (connections are always kept alive) and platform-independent message-oriented middleware (MOM). Basically, there are publishers (data sources) who publish their data in an appropriate virtual channel (topic) at a message broker entity. Data recipients then will subscribe to the appropriate topic at the broker to receive asynchronously the message through it. There are further advantageous features of MOMs that include routing and hence distributed operation (with load balancing between brokers), message queuing or support for transactions. A widely known MOM instance is the Advanced Message Queuing Protocol (AMQP) [30]. It can even be used in pair with its lightweight counterpart, the Message Queue Telemetry Transport (MQTT) [31], the popular IoT sensor node protocol. Secondly, the data representation and the message structures and types used by systems from different origin and vendors might differ completely. Therefore, establishing interactions between these systems and data aggregation from these heterogeneous sources pose a significant impediment for developers. This can be mended e.g. by the usage of the MIMOSA standard and its information metamodel. However, further engineering steps need to be made for resource constrained use cases, where the sensor has limited capability of transmitting its measurement, since the MIMOSA ontology defines a very detailed (and therefore long) descriptor for every measurement location and type. This step involves the analysis of which parts of the MIMOSA ontologies are important for the use case, and the definition of a restricted data model compatible with the full MIMOSA one, where only a data subset is taken into account. Thirdly, online machine-condition data might be available in very high resolution, and might be aggregated from multiple machines across multiple production sites. Furthermore, machine learning and statistical models can only be properly trained and created the best if there is extremely high amounts of lifetime data for various components and machines. As a consequence, big data approaches and technologies are required and can be used for CBM. It can include the online monitoring of RUL, detection of events and failures, and on-demand root cause analysis of a given failure. This can be supported by distributed stream processors that process the inbound data streams from machines and sensors. If more complex, offline asynchronous algorithms are required, then there are batch processors as well that can pull data from multiple sources (e.g. historical data from various databases). It is also possible to only generate triggers based on events detected by the stream processors. These triggers can be then distributed to the appropriate stakeholders using the message distribution system (MOM), e.g. by contacting the appropriate personnel through a human-machine interface - HMI. Such big data frameworks exist as whole solutions (e.g. offered by Microsoft Azure [32]) or even as open source as well (e.g. the Apache Spark [33] and Storm [34] frameworks). Live data can be stored in non-SQL based, highly scalable and distributed data storage systems as well (e.g. Hadoop [35] or Cassandra [36]). Finally, there can be use cases where the devices measuring and actuating the physical world can directly send their all their data to a central cloud-based backend, where all processing can be done. For these cases, this edge level can consist fully of relatively constrained systems: sensors, actuators and machines connected through a gateway. They can directly communicate with the cloud, and all data can be sent in. However, there are cases, where there can be e.g. security concerns or communication limits, which does not allow for a simple one-way data stream towards “the cloud”. For these cases (and for scalability reasons), multiple level processing is often required as mentioned in section 5. This can be solved by a “miniature” but fully-fledged processing environment that should be established right on site: all CBM related algorithms (i.e. RCA or RUL) need to run locally on dedicated hardware, in a closed and managed network. This can be done for run time, when the appropriate models have been established and deployed (i.e. production-ready machine learning algorithms well-trained for the use case). Nevertheless, since big data and statistics based analytic systems require large amount of data (possibly stemming from multiple production sites, machine types, etc.), these “local condition monitoring clouds” need to sync up parts of their data into a “higher level cloud”. The purpose of this is to feed the most possible amount of data for refinement of the models used for runtime CBM (as indicated in section 2). Currently, this “inter-cloud” synchronization (possibly on-demand, and involving partial historical data) is not solved yet, but might be supported by distributed file systems, such as Hadoop [35]. On a side note, however, the usage of these file systems as storage have not been researched for cases where primarily the MIMOSA standard database is used.

6. THE PROACTIVE MONITORING AND MAINTENANCE BUSINESS LANDSCAPE

In order to survive and thrive in a globalised market, companies are being forced to develop intelligent maintenance solutions and move towards a balanced mix of product- and client-focused approaches. Hence, companies will benefit from high service gross margins, much less adjusted than product traditional ones:

- 1) Traditional pay-per-use services are common practice, and also services that are included within the asset's price or warranty.
- 2) To this day, PMM services are not independent businesses nor have their own operating account. They are integrated within the company's global maintenance services.
- 3) Other monetisation methods (pay-per-use, availability pay, payment by results) and risk sharing between the customer and the vendor are not as usual but they are expected to become more popular.

This paper has emphasised the need for acquiring certain technology capabilities in order to be able to offer PMM products and/or services. This means investing in technical, human and financial resources. For this purpose, there are different strategies: in-house development, technology vendor acquisitions, and diverse collaboration and partnership efforts. PMM implies adding intelligence and connectivity to the end product, and it also requires to promote service based business models in contrast of traditional product centric models. The reasons behind this shift are many and varied: to meet customer needs, to leverage asset's characteristics, to implement technology innovations, etc. As stated by Brisk Insights market analysis, the global operational predictive maintenance market will grow at a CAGR (Compound Annual Growth Rate) of 26.6 % within 2016 – 2022, foreseeing a total market value of EUR2.900 million by the end of such period. This will be certainly boosted by the IIoT market rise, which is growing at a CAGR of 42 %, and will act as an enabler for its rapid industrial penetration. One of the key sectors (among all industries), in which predictive maintenance will make a huge difference will be manufacturing. The European manufacturing sector accounts for 2 million companies and 33 million jobs, representing the 15% of the total EU GDP. With the aim of increasing this contribution to 20% by 2020, European manufacturing industry faces a huge but promising challenge, given industry's potential in jobs & growth creation. However, industry's share in the EU GDP has declined during the last years, mainly due to a deceleration of global investments, market uncertainty and production offshoring to low-cost countries. This applies to all actors of the manufacturing value chain, involving production asset end users, asset manufacturers and asset service providers. In order to cope with that, the full digitisation of European industrial ecosystems has been stated as the foundation upon which competitiveness goals will be achieved. Within this framework, predictive maintenance accounts for a huge improvement potential to all actors mentioned: relevant productivity increase (asset end users), new revenue streams with higher profit margins (asset manufacturers) and new business opportunities based on analytics (asset service providers). According to McKinsey, predictive maintenance in factories could cut maintenance costs down by 10 to 40 percent, leading to manufacturer's savings of 215 to 580 billion euros in 2025, resulting from reduced downtimes and minimised manufacturing defects among others. Despite this clear potential, maintenance strategies in place still rely on ineffective corrective and preventive maintenance actions, which have a high impact on productivity (higher production costs, delays on delivery, customer dissatisfaction, etc. Not only available shop floor data and production assets' behaviour knowledge is underutilised, but also new businesses generation along the value chain is completely hampered. Regarding technology, there are several reasons behind the lack of adoption of predictive maintenance across EU industries:

- 1) Production systems complexity: the majority of EU industrial facilities is shaped by very heterogeneous assets, being the asset end user unable to gather deep knowledge about the behaviour of each asset (expertise often retained by the asset manufacturer). Heterogeneous data needs to be collected in an efficient way.
- 2) Lack of interoperability among different assets: afraid of the possibility of having a 3rd party providing services on their production assets, asset manufacturers often apply vendor lock-in solutions to their products. This results in a huge IT integration work required to connect them, usually preventing end users from implementing predictive maintenance solutions.
- 3) Non-reliable prognostics estimates at system level: even though successful prognostics applications have been deployed at component and sub-system level, asset end users' interest focuses on increasing the availability of the whole system, which has a direct impact on competitiveness. Thus, the lack of real prognostics & health management systems demonstrated at industrial level derives in a reluctance in early adopters.

In order to overcome those limiting factors, there is a clear need of bringing together all value chain actors (gathering real time data, asset behaviour knowledge and analytics expertise); as well as taking advantage of advanced analytics technologies already applied in a wide range of sectors. This will enable to match predictive management system capabilities with real industrial needs,

achieving downtime minimisation and OEE maximisation at system level. Besides all above, several non-technological challenges (such as corporate culture) prevent the penetration of predictive maintenance technologies across industries. This applies especially to SMEs, being the most relevant the following ones: 1) Uncertain Return on Investment (ROI): industrial CAPEX plans are fully subject to their expected profitability, usually in a short term (depending on the company's balance sheet, often 2-3 years). Since the implementation of such a predictive maintenance systems may imply investing in data acquisition, industrial communications & advanced analysis technologies (mainly regarding old production assets), companies often opt for more profitable investments (e-g purchasing new machinery, which leads to a direct productivity improvement). 2) Required skills: despite the high level of automation in place in most of European industrial companies, the implementation of Industry 4.0 (within which predictive maintenance is located) is currently requiring a shift from classical operator to highly analytical profiles. Industrial HMIs usually do not take advantage of available technologies such as adaptability, self-learning features, etc., resulting in workers' frustration by not showing the right information to the right people. With this in mind, it is fundamental that each company defines clearly their core business and performs a strategic planning to achieve the goal they seek. In any case, PMM opens the possibility to define and develop new business models where more companies will be involved throughout the product's value chain.

7. CONCLUSION

Proactive monitoring and maintenance enabled new business models shall address the optimized value proposition for each organisation. Personalized identification of a Canvas model element, such as key partners, key activities, key resources, customer segments, customer relationships and channels, derive a revenue stream that should definitely strength financial outcome.

ACKNOWLEDGEMENTS

The research has been conducted as a part of MANTIS Cyber Physical System based Proactive Collaborative Maintenance project. The project has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 662189. This Joint Undertaking receives support from the European Union's Horizon 2020 research and the national funding organisations Finnish Funding Agency for Innovation Tekes, Ministerio de Industria, Energía y Turismo (Spain) and by National Funds through FCT/MEC (Portuguese Foundation for Science and Technology), within project ECSEL/0004/2014, JU grant nr. 662189 (MANTIS); by the Portuguese Fundação para a Ciência e a Tecnologia (FCT, I.P.) in the framework of project UID/EEA/00066/2013 PEST (Strategic Plan for Science and Technology) for the Centre of Technology and Systems (CTS) and by the Hungarian National Research, Development and Innovation Office (NKFI), within project MANTIS, NEMZ_15-1-2016-0021.

REFERENCES

- [1] Omer, A. I., & Taleb, M. M. Architecture of Industrial Automation Systems. *European Scientific Journal*, ESJ 10.3 (2014)
- [2] Origin of OEE. OEE Foundation. Retrieved from: <http://www.oeefoundation.org/origin-of-oee/>
- [3] Understanding OEE. Retrieved from: <http://www.oeefoundation.org/origin-of-oee/>
- [4] MIMOSA. Retrieved from: <http://www.mimosa.org/mimosa/>
- [5] ISO 13374-1: Condition monitoring and diagnostics of machines - Part 1: General Guidelines
- [6] Arnaiz, A., Holmberg, K., Adgar, A., Jantunen, E., Mascolo, J. & Mekid, S. (2010), "A New Integrated E-maintenance Concept" in *E-Maintenance*, pp. 52-53.
- [7] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L. & Siegel, D. (2014). "Prognostics and health management design for rotary machinery systems reviews, methodology and applications." *Mechanical Systems and Signal Processing*, 42, 314-334.
- [8] Thurston, M. & Lebold, M. (2001) "Standards development for condition-based maintenance systems." *New frontiers in integrated diagnostics and prognostics*. 55th Meeting of the Society for Machinery Failure Prevention Technology, MFPT (2001), pp (363-373).
- [9] Lee, E. A., & Seshia, S. A. *Introduction to Embedded Systems: A Cyber-physical Systems Approach*, <http://LeeSeshia.org>, ISBN 978-0-557-70857-4, 2011
- [10] *Cyber-Physical Systems - Driving force for innovations in mobility*, Acatech, Springer. 2011.
- [11] Marwedel, P. *Embedded System Design - Embedded Systems Foundations of Cyber-Physical Systems*. 2011.
- [12] Sanislav, T. & Miclea, L. "Cyber-Physical Systems - Concept, Challenges and Research Areas," *J. Control Eng. Appl. Inform.*, vol. 14, no. 2, pp. 28-33, Jun. 2012.

- [13] Iarovyi, S., Mohammed, W. M., Lobov, A., Ferrer, B. R. & Lastra, J. L. M. "Cyber-Physical Systems for Open-Knowledge-Driven Manufacturing Execution Systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1142–1154, May 2016.
- [14] Huang, B. X. "Cyber physical systems: a survey," Jun-2008.
- [15] Engell, S., Paulen, R., Reniers, M. A., Sonntag, C. & Thompson, H. "Core Research and Innovation Areas in Cyber-Physical Systems of Systems," in *Cyber Physical Systems. Design, Modeling, and Evaluation*, 2015, pp. 40–55.
- [16] Uckelmann, D., Harrison, M. & Michahelles, F. "An Architectural Approach Towards the Future Internet of Things," in *Architecting the Internet of Things*, Eds. Springer Berlin Heidelberg, 2011, pp. 1–24.
- [17] Soldatos, J., Gusmeroli, S., Malo, P. & Di Orio, G. "Internet of Things Applications in Future Manufacturing," in *Digitising Industry - Internet of Things Connecting the Physical, Digital and Virtual Worlds*, River Publishers, 2016.
- [18] Colombo, A. *et al.*, Eds., *Industrial Cloud-Based Cyber-Physical Systems: The IMC-AESOP Approach*, 2014 edition. New York: Springer, 2014.
- [19] Leitão, P., Colombo, A. W. & Karnouskos, S. "Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges," *Comput. Ind.*, vol. 81, pp. 11–25, Sep. 2016.
- [20] Badger, L., Grance, T., Patt-Corner, R. & Voas, J. "Cloud Computing Synopsis and Recommendations."
- [21] Vollmer, K., Gilpin, M. & Rose, S. "The Forrester Wave™: Enterprise Service Bus, Q2 2011" April 25, 2011
- [22] Arrowhead -project, Grant Agreement no: 332987-2, Artemis Innovation Pilot Project, ART-010000-2013-3.
- [23] Bonomi, F., Milito, F., Zhu, J. & Addepalli, S. "Fog Computing and Its Role in the Internet of Things," in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, New York, NY, USA, 2012, pp. 13–16.
- [24] LaMothe, R. "Edge Computing." Pacific Northwest National Laboratory, Jan-2013.
- [25] Gustafsson, A. *et al.*, "The relevance of service in European manufacturing industries," *J. Serv. Manag.*, vol. 21, no. 5, pp. 715–726, 2010.
- [26] Di Orio, G., Matei, O., Scholze, S., Stokic, D., Barata, J. & Cenedese, C. "A Platform to Support the Product Servitization," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 7, no. 2, 2016.
- [27] Vandermerwe, S. & Rada, J. "Servitization of business: Adding value by adding services," *Eur. Manag. J.*, vol. 6, no. 4, pp. 314–324, 1988.
- [28] Roy, R. *et al.*, "The servitization of manufacturing: A review of literature and reflection on future challenges," *J. Manuf. Technol. Manag.*, vol. 20, no. 5, pp. 547–567, 2009.
- [29] Lee, J. & Bagheri, B. "Cyber-Physical Systems in Future Maintenance," in *9th WCEAM Research Papers*, Springer, Cham, 2015, pp. 299–305
- [30] Advanced Message Queuing Protocol (AMQP). Retrieved from <https://www.amqp.org/>
- [31] Message Queue Telemetry Transport (MQTT). Retrieved from <http://mqtt.org/>
- [32] Microsoft Azure. Retrieved from https://azure.microsoft.com/en-gb/?wt.mc_id=AID539500_SEM
- [33] Apache Spark. Retrieved from <http://spark.apache.org/>
- [34] Apache Storm. Retrieved from <http://storm.apache.org/>
- [35] Apache Hadoop. Retrieved from <http://hadoop.apache.org/>
- [36] Apache Cassandra. Retrieved from <http://cassandra.apache.org/>

Nuisance Attribute Projection Based Channel Compensation Technique and its Application in Bearing Performance Degradation Assessment

Huiming Jiang¹, Jin Chen¹, Guangming Dong¹

¹ State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, No.800, Dongchuan Rd, Shanghai 200240, China

ABSTRACT

During condition monitoring, the sampled data covers not only information about the health state or fault degree, but also information about the sample channel. The information about the sample channel is nuisance attribute for the condition monitoring. However, it induces different information for different channels, which influences the performance of the condition monitoring system. In this paper, the nuisance attribute projection (NAP) based channel compensation technique is introduced. NAP can filter out the effect of nuisance attributes in feature space through projection. Performance degradation assessment method based on hidden Markov model and NAP is proposed. The results of experimental data show that NAP can effectively improve the accuracy and robustness of the performance degradation assessment system.

Keywords: Nuisance attribute projection, Channel compensation, Performance degradation assessment, Hidden Markov model, bearing, Projection

Corresponding author: Guangming Dong (gmdong@sjtu.edu.cn)

1. INTRODUCTION

For machine learning based monitoring system, the feature extracted from the sampling data affects the final performance of the system directly. In order to improve the effectiveness of the feature, many feature transformation methods have been developed, such as principal component analysis (PCA)[1], kernel principal component analysis (KPCA)[2] and locality preserving projections (LPP)[3] and so on. These feature transformation techniques try to save the most important information into lower dimension space, and try to improve the effectiveness at the same time. However, the sampled data covers not only information about the health state or fault degree, but also information about the sample channel during the condition monitoring. The variability of the channel and environment is one of the most important factors affecting the performance of final systems[4]. Furthermore, the difference between the sample channel of the training and testing data can cause the relevant performance degradations. And less research has been carried out on this problem. Nuisance attribute projection (NAP) is an effective technique which can filter out the effect of nuisance attributes in feature space [5]. NAP was initially introduced to mitigate the effect of the channel variability in the speaker recognition field [6, 7], face recognition and image recognition [5, 8]. In this paper, we introduce NAP as a channel compensation technique into the bearing performance degradation assessment system.

Bearings are one of the most widely used elements in rotary machines and their failures are one of the most frequent reasons for machine breakdown[9]. Bearing performance degradation assessment has attracted many attentions from researchers [10]. Hidden Markov model (HMM) is a statistical model based on time series, which is a very suitable for performance degradation assessment [11]. As a data-driven method, the HMM based method uses features as inputs, which means the effectiveness of the extracted feature affects the final performance directly. Facing the channel interference problem, a bearing performance degradation

assessment method based on NAP and HMM is proposed in this paper. NAP can improve the effectiveness of the feature space and solve the mismatch of the sample channel of training and testing data effectively.

The rest of the paper is organized as follows: section 2 introduces backgrounds of hidden Markov model and nuisance attribute projection; section 3 describes the proposed bearing performance degradation assessment scheme based on NAP and HMM. The effectiveness of the proposed scheme is verified through a bearing accelerated life test in section 4. Then in section 5, a conclusion is drawn.

2. BACKGROUNDS

2.1. Hidden Markov model

The HMM concept is introduced in details in references [12, 13]. One HMM can be described by the following five parameters:

(1) N : the number of states in the model. Suppose S_1, S_2, \dots, S_N are N states of the model, and q_t stands for the state the model stays at the moment t . Obviously, $q_t \in \{S_1, S_2, \dots, S_N\}$.

(2) M : the number of distinct observation symbols per state. Suppose v_1, v_2, \dots, v_M are M distinct observations, and o_t stands for the observation at the moment t . Obviously, $o_t \in \{v_1, v_2, \dots, v_M\}$.

(3) π : the initial state probability distribution. $\pi = \{\pi_i\}$, where $\pi_i = P(q_1 = S_i), 1 \leq i \leq N$.

(4) A : the state transition probability matrix. $A = \{a_{ij}\}$, where $a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N$.

(5) B : the observation symbol probability matrix. $B = \{b_j(k)\}$, where

$$b_j(k) = P(o_t = v_k | q_t = S_j), 1 \leq j \leq N, 1 \leq k \leq M.$$

For convenience, the notation $\lambda = (\pi, A, B)$ is used to indicate the complete parameter set of the model. Continuous HMM is an extension of HMM for continuous signals and utilizes continuous observation densities [13]. In this paper, all HMM models used are continuous HMM without special instructions.

2.2. Nuisance Attribute Projection

In this paper, NAP is introduced into the HMM based bearing performance degradation assessment. The feature space is projected into a new space based on NAP before putting into HMM. The new feature space covers more dimensions about the statistic information of degradation stage or fault level and less dimensions corresponding to the channel nuisance attributes which have no help in describing degradation statistics.

The working of the NAP algorithm is pictorially explained in figure 1, in which m represents feature vectors with system information (containing the sample channel and the manufacturing error variations), and m^s is the system independent feature vector. The sub-space representing redundant information (such as sample channels and manufacturing errors) contains little information about the degradation stage or fault level of bearings, and we try to build a projection to zero that out.

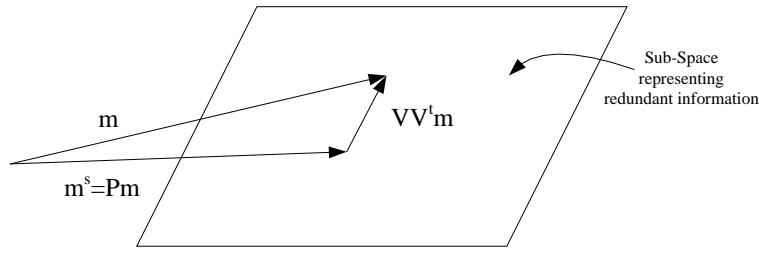


Figure 1. Nuisance attribute projection

In our application with NAP for feature space projection, it aims to find a projection matrix $\mathbf{P} = \mathbf{I} - \mathbf{V}\mathbf{V}^t$, filtering out the nuisance attributes (e.g. the channel variability and the manufacturing error variability) from the feature vectors. The matrix \mathbf{V} used in the NAP technique is found by minimizing the effect of the nuisance attributes on the feature space:

$$\delta = \sum_{i,j} \mathbf{W}_{ij} \|\mathbf{P} \cdot f_i - \mathbf{P} \cdot f_j\|^2 \quad (17)$$

where $\{f_i; i=1,2,\dots,n\}$ are n feature vectors of the testing bearing, and \mathbf{W} is a weight matrix, whose entries correspond to pairs of feature vectors. \mathbf{W}_{ij} is positive for pairs of vectors that we want to move together, negative for pairs we do not move together, and zero for pairs we do not care if they come together or not. In our applications, feature vectors extracted from the normal condition data, but from different channels, need to be moved together. Feature vectors from the same channel need to be maintained. So the weight matrix for different bearings compensation can be expressed as:

$$\mathbf{W}_{ij} = \begin{cases} 1; & \text{when } channel(f_i) \neq channel(f_j) \\ 0; & \text{Otherwise} \end{cases} \quad (18)$$

The objective function in Equation (1) can be minimized by k eigenvectors with largest eigenvalues of the symmetric eigenvalues problem [5]:

$$\mathbf{A}\mathbf{Z}(\mathbf{W})\mathbf{A}^t \mathbf{v}_k = \mathbf{v}_k \Lambda \quad (19)$$

where the matrix $\mathbf{Z}(\mathbf{W}) = \text{diag}(\mathbf{W} \cdot \mathbf{I}) - \mathbf{W}$, \mathbf{I} is the column vector of all ones. $\text{diag}(\cdot)$ is an operator of forming a diagonal matrix from a vector, and $\mathbf{A} = [f_1, f_2, \dots, f_n]$. Then a projection matrix \mathbf{P} can be obtained. \mathbf{P} can be utilized for the feature space projection in the preprocessing procedure for the HMM based bearing performance degradation assessment.

3. PERFORMANCE DEGRADATION ASSESSMENT BASED ON NAP AND HMM

Performance degradation assessment based on HMM covers two parts, which are normal HMM training and performance evaluation. During the training part, normal condition data are used to train normal HMM. Then during the evaluation part, the testing data is putted into the normal HMM with a likelihood probability as output. The likelihood probability indicates how well the feature vector matches the probability distribution of HMM trained by the normal data [14]. If the feature comes from the normal state, the likelihood probability tends to be around the usual values of the training data because it follows the distribution perfectly. On the other hand, if the condition to be evaluation degenerates away from the normal condition, the likelihood probability tends to be smaller than usual. Therefore, the likelihood probability can be chosen as the performance index (PI). The trend of PI indicates the degradation process of the testing bearing. This is the core idea of the HMM based performance degradation assessment.

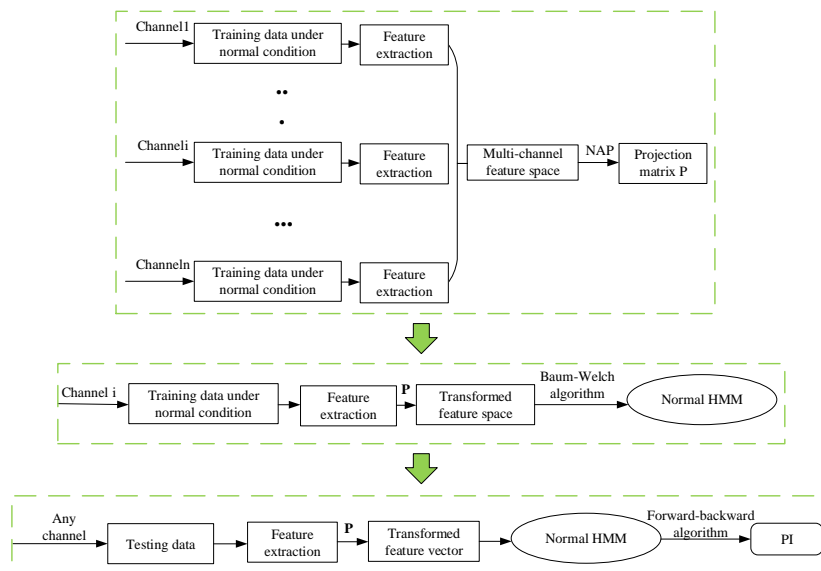


Figure 2. Procedures of the performance degradation assessment based on NAP and HMM

In the performance degradation assessment method based on NAP and HMM, NAP works as a preprocessing for HMM to provide enough and effective features. So the proposed method covers three main procedures, which are projection matrix calculation, HMM training and performance evaluation. The main procedures of the performance degradation assessment method based on NAP and HMM can be described as figure 2.

4. EXPERIMENT VERIFICATION

The bearing performance degradation assessment method based on NAP and HMM is applied to an accelerated bearing experiment to verify its effectiveness.

4.1. Experiment introduction

The accelerated bearing life tester (ABLT-1A) is provided by Hangzhou Bearing Test & Research Center (HBRC), which is shown in figure 3. The tester includes an AC motor, a transmission system, a lubrication system, a loading system and a data acquisition system. Four bearings are installed on the test rig for the accelerated bearing life test. The sketch of load and the location of sensors are shown in figure 4. The test conditions are shown in table 1. The parameters and characteristic frequencies of testing bearings are shown in table 2 and table 3.

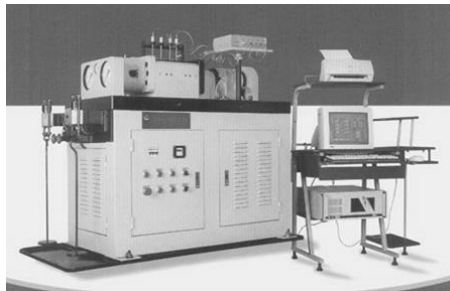


Figure 3. Accelerated bearing life tester (ABLT-1A)

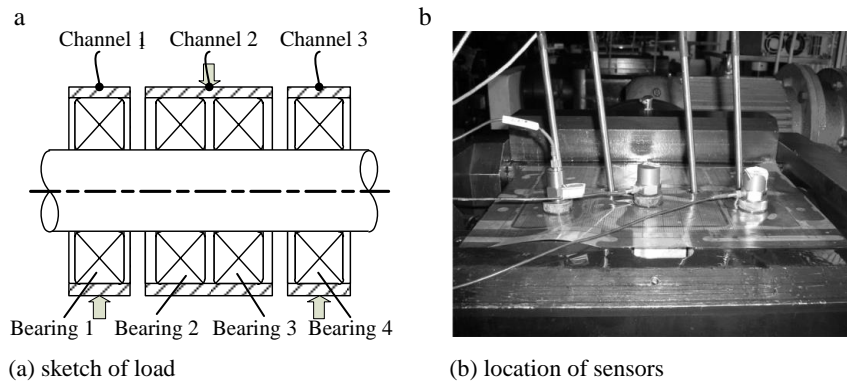


Figure 4. Sketch of load and location of sensors

Table 1: Test conditions

| Testing conditions | Detailed parameters |
|--------------------|------------------------------|
| Bearing type | Rolling element bearing 6307 |
| Number of bearings | 8 |
| Load | 12.744kN |
| Rotate speed | 3000rpm |
| Rating lifetime | 9000min |
| Sample frequency | 25.6kHz |
| Sample period | 0.8s |
| Sample interval | 1min |

Table 2: Rolling element bearing parameters

| Type | Ball number | Ball diameter (mm) | Pitch diameter (mm) | Contact angle | Motor speed (rmp) |
|------|-------------|--------------------|---------------------|---------------|-------------------|
| 6307 | 8 | 13.494 | 58.5 | 0 | 3000 |

Table 3: Five characteristic frequencies of rolling element bearing 6307(Hz)

| f_r | f_c | f_b | f_i | f_o |
|-------|-------|-------|-------|-------|
| 50 | 19 | 102 | 246 | 153 |

4.2. Data analysis

During the experiment, B12 on the location of Bearing 1 in figure 4 run to failure at first when other bearings are under normal condition. So the sampled three channel data cover most information about the degradation process of B12. The data of B12 is analyzed in this part. The whole life time of B12 is 2469min. The time plots and RMS curves of three channels are displayed in figure 5. From figure 5, it is obviously that the vibration data and RMSs of three channels are different from each other. The data in each channel covers not only the information about the degradation stage of the bearing but also interference information introduced by the sample channel, which affects the final performance of the HMM based performance degradation assessment. In this paper, NAP is used to move out these nuisance attributes caused by channel variabilities.

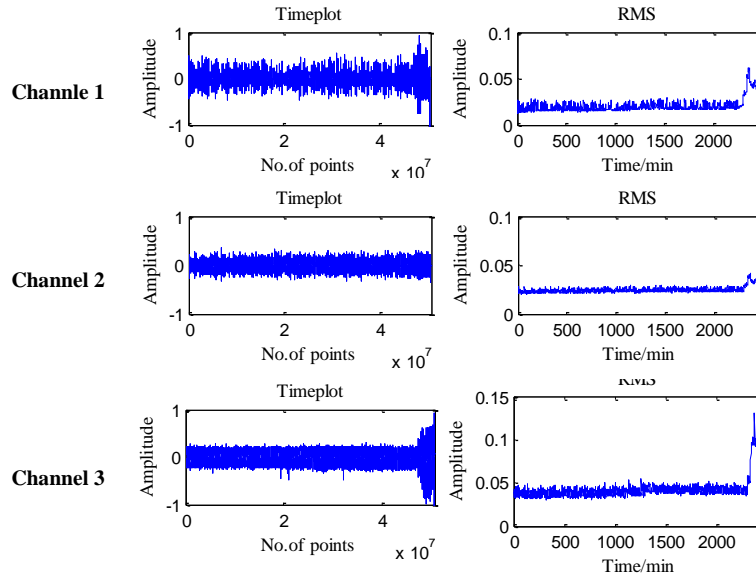


Figure 5. Time plots and RMS of B12

4.3. NAP matrix calculation

29-dimension feature vector is extracted for each sample, which covers eleven time domain features, eight wavelet packet decomposition node energies and eight energy entropies when level is 3, amplitude spectrum entropy and envelope spectrum entropy. Feature vectors of three channels during the first 10 minutes are used to calculate the projected matrix. They cover the same fault information and different channel attributes. The weight matrix is set as follows:

$$\mathbf{W}_{ij} = \begin{cases} 1 & \text{if } channel(f_i) \neq channel(f_j) \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

Through NAP calculation by Equation(19), a matrix \mathbf{P} is obtained. To show the effectiveness of NAP, the feature vectors before NAP and after NAP are displayed in figure 6. Differences caused by channel variabilities are projected out obviously.

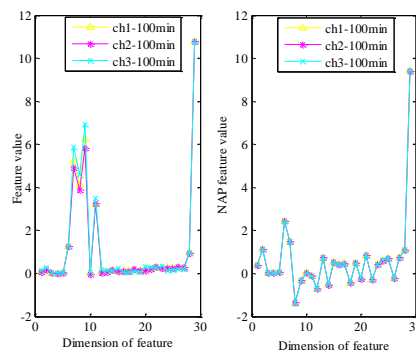


Figure 6. Feature vectors before NAP and after NAP

4.4. HMM training

From figure 5, it can be inferred that B12 is under normal condition during the first 500 minutes. The normal condition data before 500 min is used to train normal HMM. Before NAP can remove nuisance attributes caused by channel in data. Any channel data can be used to train the normal HMM. Channel 3 is selected as an example. Features are transformed by matrix \mathbf{P} calculated in section 4.3 before put into HMM. A normal HMM is obtained finally.

4.5. Performance evaluation

The whole life time data of channel 3 is used to evaluate the degradation process. After feature extraction, feature vectors are transformed by matrix P calculated in section 4.3 before put into the normal HMM. The output PIs are used to evaluate the degradation process as shown in figure 7. The whole life PI covers three obvious stages. Before 1295min, PI keeps stable with little fluctuations. After 1295min, PI drops into a lower level with bigger fluctuations. And this stage keeps for a long time. Then PI increases a little at 2307min and drops suddenly at 2337min until failure. It is inferred that B12 is under normal condition before 1295min and under early degradation since 1295min. Then early fault occurs at 2307min, and the fault gets worse at 2337min until final failure. To verify these conclusions, the time plots and envelop spectra of key points are displayed in figure 8. And the final fault of B12 is inner race fault as shown in figure 9. Figure 8 and figure 9 verify the effectiveness of the performance degradation assessment method based on NAP and HMM.

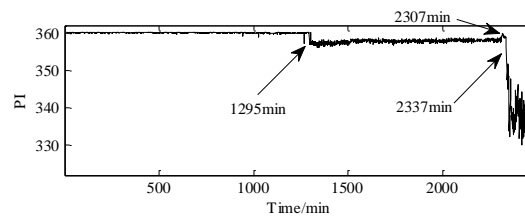


Figure 7. the PI curve

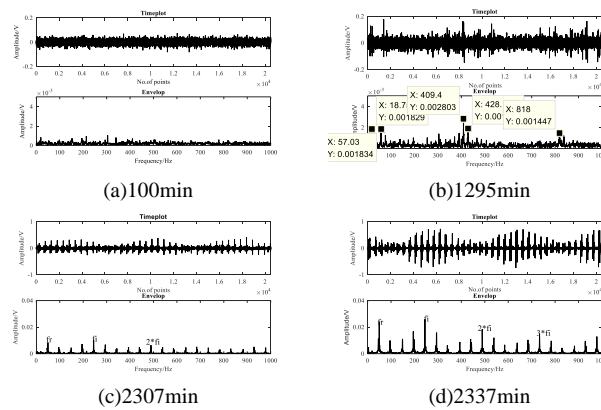


Figure 8. Time plots and envelop spectra of key points



Figure 9. the final failure

After NAP, the feature vector covers little information about the channel. So the normal HMM trained by channel 3 can be used to assess data sampled by other channels. The evaluation results by other channel testing data are shown in figure 10.

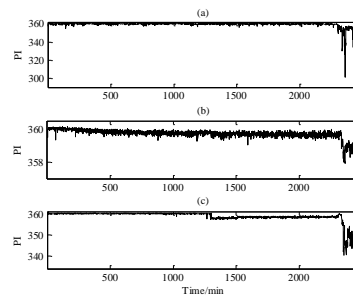


Figure 10. (a)PI of channel 1 (b)PI of channel 2 (c)PI of channel 3

In order to show the importance of NAP in this method, the performance degradation assessment method without NAP is carried out for comparison. Final results are shown in figure 11 and figure 12. Comparing figure 7 and figure 11, it is obvious that NAP can improve the representational capacity of features especially for the early fault. Comparing figure 10 and figure 12, it is verified that NAP can remove the nuisance attributes caused by channel variabilities and improve the robustness of HMM which solves the problem caused by the mismatch of the sample channel of training and testing data effectively. In conclusion, the bearing performance degradation assessment method based on NAP and HMM is effective.

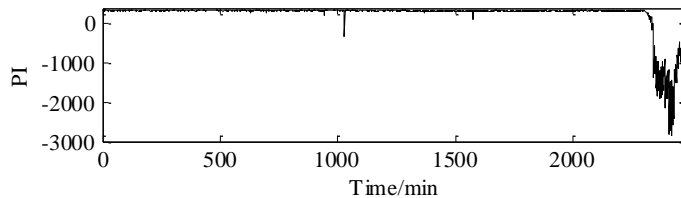


Figure 11. PI of channel 3 through the method without NAP

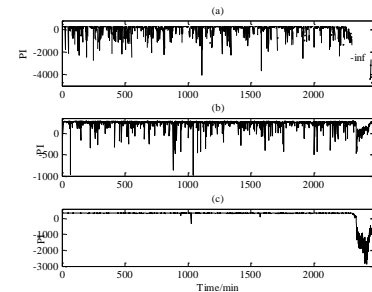


Figure 12. PIs through the method without NAP:(a)PI of channel 1 (b)PI of channel 2 (c)PI of channel 3

5. CONCLUSIONS

Interferences caused channel variabilities is urgent to be solved and important for bearing performance degradation assessment. In this paper, Nuisance attribute projection is introduced as a channel compensation technique and is combined with HMM for bearing performance degradation assessment. The performance degradation assessment based on NAP and HMM is proposed. NAP can effectively mitigate these influences irrelevant to the degradation state which are caused by channel attributes. An accelerated bearing experiment is used to verify the effectiveness of the proposed method. The results of experimental data show that NAP can effectively improve the accuracy and robustness of the performance degradation assessment system.

REFERENCES

- [1] Malhi, A., Gao, R. X. (2005) PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation & Measurement*. 53(6): 1517-1525.
- [2] Kim, K. I., Jung, K., Hang, J. K. (2002) Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*. 9(2): 40-42.
- [3] Dornaika, F., Assoum, A. (2013) Enhanced and parameterless Locality Preserving Projections for face recognition. *Neurocomputing*. 99(1): 448-457.
- [4] Vair, C., Colibro, D., Castaldo, F., Dalmasso, E., Channel Factors Compensation in Model and Feature Domain for Speaker Recognition, in: *Speaker and Language Recognition Workshop*, 2006. *IEEE Odyssey 2006: The*, 2006, 1-6.
- [5] Solomonoff, A., Campbell, W. M., Quillen, C. (2007) Nuisance attribute projection. *Speech Communication*.

- [6] Campbell, W. M., Sturim, D. E., Reynolds, D. A., Solomonoff, A., SVM based speaker verification using a GMM supervector kernel and NAP variability compensation, in: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, IEEE, 2006, 1-1.
- [7] Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P. (2011) Front-end factor analysis for speaker verification. *Audio, Speech, and Language Processing, IEEE Transactions on.* 19(4): 788-798.
- [8] Struc, V., Vesnicer, B., Mihelic, F., Pavesic, N., Ieee (2010) REMOVING ILLUMINATION ARTIFACTS FROM FACE IMAGES USING THE NUISANCE ATTRIBUTE PROJECTION. 2010 Ieee International Conference on Acoustics, Speech, and Signal Processing. 846-849.
- [9] Randall, R. B., Antoni, J. (2011) Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing.* 25(2): 485-520.
- [10] Tran, V. T., Hong, T. P., Yang, B. S., Tan, T. N. (2012) Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mechanical Systems & Signal Processing.* 32(4): 320-330.
- [11] Jardine, A. K., Lin, D., Banjevic, D. (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing.* 20(7): 1483-1510.
- [12] Rabiner, L. R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE.* 77(2): 257-286.
- [13] Jiang, H., Chen, J., Dong, G., Liu, T., Chen, G. (2015) Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis. *Mechanical systems and signal processing.* 52(338-359).
- [14] Ning, C., Chen, M., Zhou, D. (2014) Hidden Markov Model-Based Statistics Pattern Analysis for Multimode Process Monitoring: An Index-Switching Scheme. *Industrial & Engineering Chemistry Research.* 53(27): 11084-11095.

Dynamic State Recognition Using CNN-RNN Processing Pipeline

E. R. Anas, B. J. Matuszewski

Computer Vision and Machine Learning Research Group, University of Central Lancashire, Preston PR1 2HE, UK

ABSTRACT

The paper describes a novel method for detecting motion using a software pipeline model consisting of Convolutional Neural Network ‘CNN’ followed by Recurrent Neural Network ‘RNN’. The method can be useful to monitor a sequence of events and provide a decision of action completion. The CNN, generally, provides a sufficient information about the object states but without information describing temporal dependence between these states. By introducing an RNN network after the CNN, the software could utilize the temporal information about the system present in the successive observations. To prove the concept, the human’s eye blink action is used as an example. The training procedure consists of a CNN first trained on eye images to detect the distinctive eye states. Subsequently, the eye status probability values predicted from the trained CNN divided into short sequences contain the temporal information describing the blink action. Finally, a multilayer RNN has been trained on these sequences to detect the eye blink events. The adopted RNN model shows a capability to capture relevant temporal information even when different CNNs, trained on the same datasets, are used. More specifically, the proposed method has achieved 93%/94% and 98%/98% recall/ precision on the Eyeblink8 and Talking Face datasets respectively.

Keywords: Motion detection, Monitoring, Deep Learning, Eye Status Detection, Eye Blinking Rate Estimation.

Corresponding author: Essa Anas (eana@uclan.ac.uk)

1. INTRODUCTION

Spatio-temporal representations describe a system behaviour and interactions jointly in space and time domains. Such system characterisations are often of interest to scientific communities, as they can describe complex dynamic behaviours of the physical systems. Many actions that are performed by humans or machines can be modelled by spatio-temporal interactions performed in a sequential fashion. One of the applications, for which such modelling is very useful, is detection, description and quantification of spatially distributed motion. Motion has an important implementation in different artificial intelligence fields. Applications like robot vision, game control, intelligent video surveillance, human-computer interaction are just a few examples for which spatio-temporal motion modelling could be of interest.

In this work, the aim is to provide evidence that the output of the CNN classification model can contain temporal information useful for a spatio-temporal description of a system/object. If the classification predictions are sequenced in time, a multilayer RNN can extract information that could be useful for estimation/detection of spatially distributed motion. Similar processing pipelines were previously proposed for image captioning, by utilizing the output of the CNN classifier to generate natural language descriptions of images. However, to our knowledge, such processing pipeline hasn’t been used for spatially distributed motion quantification. In this study, the eye blink action is used as an example to prove the concept. First, the CNN network has been trained for eye status classification as in [1]. Then, videos of subjects blinking have been applied on the CNN to collect sequences that can be used to train a multilayer RNN. In the final stage, a multilayer RNN is used to detect the blinking. This study also shows that the RNN is not strongly dependent on the specific type of the CNN used in the processing pipeline. That is, despite different possible CNN implementations, the RNN network can correctly recognise the dynamics of the system represented in the observed sequence.

The rest of the paper is organized as follows. Section 2 contains a brief review of the relevant CNN-RNN implementations followed by a report on the latest eye blink detection studies. The eye blink detector requirements are discussed in section 3. In section 4, the training of the CNN and RNN are both discussed with the system operation details and the datasets used for the system validation. The results obtained and the conclusions are reported in Sections 5 and 6, respectively.

2. RELATED WORK

It had been mentioned that CNN-RNN pipeline has been implemented for image captioning in [2]. They used a pretrained CNN to provide the classification about the objects and scenes in the image and then a single hidden layer bidirectional RNN that been trained on natural language is attached to the CNN to describe the applied image.

In the field of eye blink detection, an application like motion vectors has been employed for blink detection. The authors in [3] used Lucas-Kanade tracker (LKT) for that purpose. Viola-Jones detector first performed for the face/eye region detection. The eye region is then subdivided into 3x3 cells and 9 motion vectors are estimated as an average for each of this area. Observing the average values of these vectors, the upper 6 vectors gave a clear indication of the eyelid motion. The variance of the motion vectors that are related to the eyelid motion had been compared to an empirically selected threshold value to evaluate the blink action. The implementation of the solution included a state machine to perform the status part of the detection process. They method achieved a precision of 91% and a recall of 73.1% on the ZJU dataset. The authors also developed their dataset called Eyeblink8 and they provided it on the internet. On this dataset, the showed a 79% precision and 85.27% recall.

Modification suggested by [4] to amend the use of the approach reported in [3] that could perform better. The modification included the usage of normalized motion vectors that distributed evenly in the motion area. They estimated the motion vector at each pixel within the detected eye area using the Farneback algorithm. The proposition is that a linear relation between the intraocular distance and the eye region size can be employed to normalize the motion vectors. The normalization helped to minimize the effects of other movements, like a head motion on the estimation of the eye blink. A similar state machine, as in [3], was adapted to estimate the blinking. The reported results showed, a 100% and 98.08% for precision and recall on the ZJU dataset, 94.7% and 99% on the Eyeblink8 dataset, 95% 93.44% for Talking Face dataset. The authors also developed their own dataset coined "Researcher's Night" which is not yet available online and reported 92.42% and 81.48% for precision and recall, respectively.

Researchers approached the blink detection action in two main approaches. They either identifying the eye state in each frame searching for the sequence of interest. Or by processing multiple frames analysing the motion of the eyelids. In this study, our focus is on the eye blink action as a spatio-temporal event. This action can be defined as continuous changes in the state of the eye between opened, partially opened, and closed and back to open or to partially opened state. A complete blink occurs when these sequences of states completed within a typical timeframe between 100 and 400 milliseconds. In some cases, the sequence may not be complete as described above, like the case of the incomplete blink and the extended blinks, when either the closed state does not appear or when the blink action takes a longer time to be executed [5]. The suggested model in this work trained on these events and can response effectively to these details.

3. EYE BLINK DETECTION CONSIDERATIONS

For real-world applications, a successful computer implementation of a recognition task, that human often can perform without an effort in an instant, could be very challenging. For example, detecting the eye status can be affected by the lighting condition, eye's shape, head's pose or eye's gaze leading to estimation errors. It is difficult to design an eye blink detection system that would perform equally well under different conditions [4]. Several authors tried to implement their algorithms for various circumstances [6] or impose a specific restriction on distance, lighting etc. for their algorithms to work [7-8]. Others implement their systems to be more robust to lighting conditions with half face appearance [9].

3.1. Requirement

Generally, from a perspective of the practical system applicability, the eye status detection system should:

- adapt to changing illumination conditions,
- cope with varied distance from the camera,
- be invariant, within a limit, to head pose changes,
- be resilient to a motion blur [10],
- be capable to operate in real-time,
- cope with the current standard specification of camera frame rate, around 30 fps, i.e. when a full “blinking transition” of “open” – “partially open” – “close” – “partially open” – “open” states can be detected.

3.2. Eye Blink Analysis

It had been reported [4] that an endogenous blink could last for about 0.4s. This means that for a camera operating at 30 fps, the blink action can appear within $30 \times 0.4 = 12$ frames. One of the design choices is to select the number of frames that contain the required blink sequence that could be utilized as a training set for the Recurrent Neural Network (RNN). Sequences consisting of 10, 7 and 6 frames were collected for training purposes. It was concluded that sequences of 7 frames (corresponding to about 250ms time window) provided enough contextual temporal information for the RNN to perform well and enabled efficient training at the same time.

4. CNN AND RNN TRAINING

To build a dataset for the eye status detection that can support a system that can address the issues highlighted in the previous section, the publicly available Helen dataset [12] has been utilized. This database was designed primarily for facial feature extraction in the wild. Helen dataset contains many images with subjects with opened and partially-opened eyes, however, the number of subjects with closed eyes is rather limited. Therefore, additional examples (both synthetic and real) representing “closed” eye class were added.

The Helen database consists of facial images representing subjects of different age, gender, and ethnic origin. Additionally, the images are of different resolution and were captured at highly variable illumination and pose conditions. There are 2330 images available in this dataset. From these, more than 2000 images of right and left eyes have been cropped. The same images have been duplicated with image enhancement for augmentation and their grey scale versions have been obtained to have an option to train different models, with colour or grey scale images for performance comparison. For practical reasons, each eye in the image is treated independently rather than the pair, as in some cases detecting simultaneously two eyes may fail. For instance, for rotated face or complex unbalanced lighting conditions, only one eye might be visible.

4.1. CNN Training

Two CNN models had been trained on the mentioned Helen dataset, AlexNet [11] and a concise network (named here as ConcisedNet) which was previously proposed in [1], see figure 1.

Regarding the eye status specification, three states had been defined as shown in table 1. The CNN models had been trained with these three classes.

Table 1: Classification of the eye status based on the iris visibility [1].

| Eye Status (Classes) | Iris Visibility % |
|----------------------|-------------------|
| Opened | ~100-50% |
| partially-opened | ~50-5% |
| closed | ~5-0% |

AlexNet is a well-known model. It had won the Imagenet competition in 2012, and since then, many applications had been inspired by that network. The network contains too many parameters to make it suitable for the online applications but the implementations using GPU. The ConcisedNet has a smaller number of parameters and therefore can run efficiently on a CPU. For the results reported in this paper, the two models were trained on the same dataset.

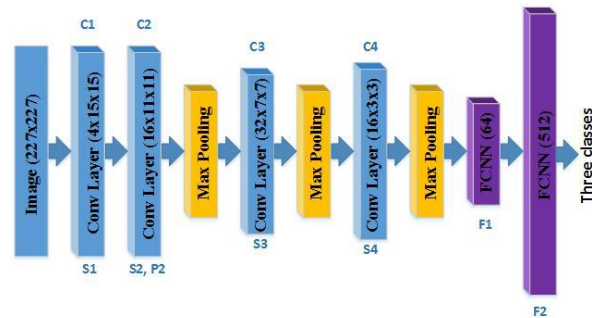


Figure 1. ConcisedNet architecture [1]

4.2. RNN Training

RNN represents a special category of neural networks that has the capability to model the sequential dynamics of the input data. This capability comes from sharing the model parameters through the temporal domain. Through training, the model can learn and generalize temporal features from the applied training set. Consequently, sharing temporal parameters allows the network to learn patterns that appear in any place or multiple places within the sequence.

The blink action can be modeled as a sequence of eye states, for instance: open, partially open, close, partially open, and open. Then, it is the matter of detecting this sequence to decide whether a blink action has taken place. However, the problem is not as simple as it seems. In most of the blink sequences examined, the sequences are not as clear as given above and thus a simple decision on the eye status could not be enough for the RNN to be trained on. In fact, using the eye status to train the RNN model provide unsatisfactory detection results. This poor performance is mainly due to the marginalization of the information about the level of the eye-opening that contain evidence about the blink action. Using the state of the eye as a feature turned to lose the temporal information between the consecutive frames and ultimately there isn't enough information left to train the sequential model.

To cope with this issue, the output of the CNN has been analysed. In Table 2, the outputs from the last layer of the CNN, the softmax layer, for images representing the same class of 'partially open' are shown. The level of confidence about different states for the given image can also be informative. For instance, in the first case (on the left) the eye is almost open while the second case (on the right right), the eye is almost closed. In this case, the confidence level about the "open" and "close" states could be used to differentiate between these two images, representing the same eye status.

Table 2: The CNN predictions, for the two images shown, are given as 'partially opened'. However, as the eyelid is closing the 'close' status probability (confidence level) is increasing. Similarly, when the eyelid is opening the 'open' status probability is increasing, i.e. getting more significant in the prediction state. The use of the state probabilities provides more meaningful information about the importance of each state when combined with results from the neighboring frames.

| | | | |
|--------------|--------|--------------|--------|
| | | | |
| partial open | 78.25% | partial open | 91.67% |
| open | 17.78% | close | 7.51% |
| close | 3.97% | open | 0.82% |

Generally, the relationship of between the probabilities regarding the softmax function can be depicted as below.

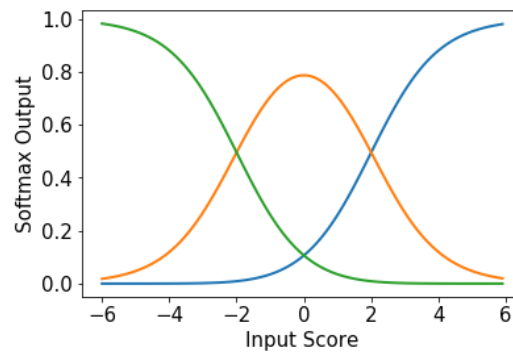


Figure 2. The softmax output represent the probability relationship between different classes.

As the score values increase it approaches 1, while when the value is low the distribution of the probability will tend to a uniform shape.

4.3. System Operation

The system operation includes the necessary software to detect the face and the eye area extracting the eye area and resize it to fit to the CNN network input. The program passes the images of the left and the right eye and obtains two readings that make a vector of six elements (3 states confidence levels for each eye). The range of the CNN output is between $[0, 1]$ since the final CNN layer, in both AlexNet and ConcisedNet, is the softmax layer. The six elements output vector for each frame is concatenated with the vectors for the previous 6 frames to build a sequence of 42 elements which are applied as an input to the RNN. In case that the eye detector is not able to detect the eye area, the CNN will not receive the required input and therefore the relevant frames are ignored. The input sequence is buffered and shifted each time the data from a new frame arrives. The input sequence contains data from the last seven frames.

To show that the RNN network is not too strongly dependent on the type of the CNN network that generates its training sequence, two different CNNs (AlexNET and ConcisedNet) were used in two separate experiments. The ConcisedNet is a network with a small number of parameters compared to AlexNET. Using smaller network could help to implement the solution on a CPU when a GPU adapter is not an option. In our solution, the RNN was trained with the target value of $\{1, -1\}$. The value of $\{1\}$ is related to a true blink sequence, while the value of $\{-1\}$ is related to a no blinking state.

4.4. Datasets

Different dataset were investigated to provide more insight into properties of the proposed solution. Each dataset has different characteristics that can affect the performance in a different way. Furthermore, subjects within the datasets exhibit different blinking characteristics. Furthermore, the same subject may blink differently within the same video. For instance, sometimes a subject performs a complete blink cycle, in other cases one eye is totally closed while the other is partially open. Or when both eyes are partially opened (or partially closed). This unsynchronized behavior makes the blink detection a harder task. This was another reason why using simply the eye state to detect eye blink action was not fully satisfactory. In this study, Eyblink8 and Talking face are used to test the proposed solution.

5. RESULTS

In the proposed method, the output from the RNN has values approximately between $[-1, 1]$. Therefore, it is necessary to provide a threshold that can be used to decide whether the output of the RNN represents a blink or not. figure 3 shows the RNN instantaneous response compared to the ground truth on part of Eyblink8 dataset. The network response shows a high margin between the true blink and the no-blink. In fact, the network output starts to appear as the blink is starting to develop and reaches its peak when the

network receives a complete pattern of the blink sequence. In some cases, the annotation is not precise so the blink action is detected from the RNN before or after the ground truth confirms the blink.

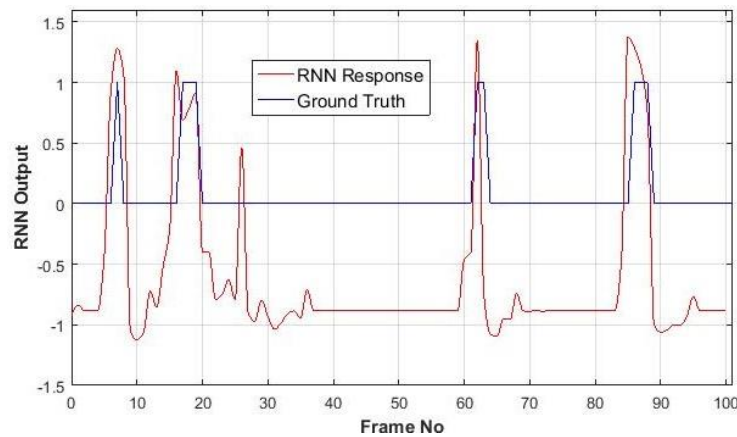


Figure 3. RNN response to the input from the Talking Face dataset.

The figures 4, 5, and 6 show the precision vs recall for two datasets and the AlexNET-RNN and ConcisedNet-RNN configurations. It is worth to mention here that although the RNN training data was obtained from the ConcisedNet CNN and only three Eyeblink8 videos, the RNN generalised well to another dataset and work also with the output from AlexNET. Although the RNN network had been trained on a sequence generated from the ConcisedNet, attaching the RNN to the AlexNET provided better results than the ConcisedNet-RNN configuration.

The three graphs show the parametric Precision versus Recall curves with the threshold value as a parameter. Using the F1 score, the estimated optimal threshold value (maximising the F1 score) and the corresponding Precision and Recall are provided.

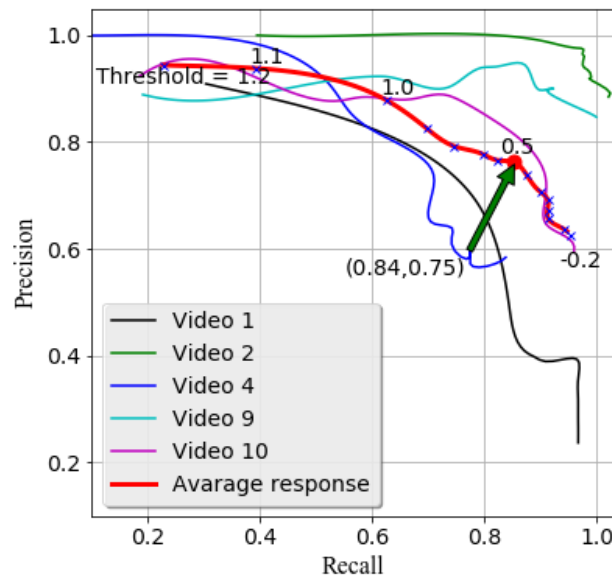


Figure 4. Using ConcisedNet, Precision vs Recall for different values of the threshold for EyeBlink8 dataset are shown. F1 score shows that Threshold = 1 provides the best results.

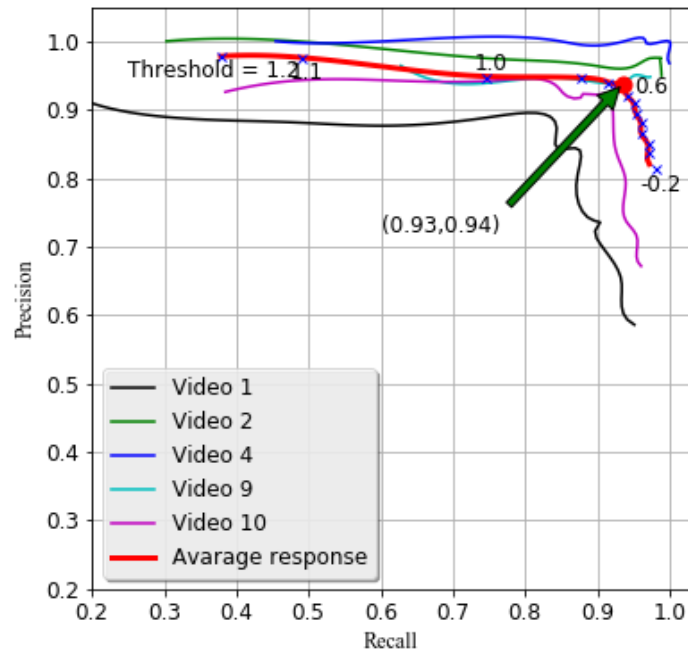


Figure 5. Using AlexNet for EyeBlink8 with highest F1 score at Threshold = 0.6.

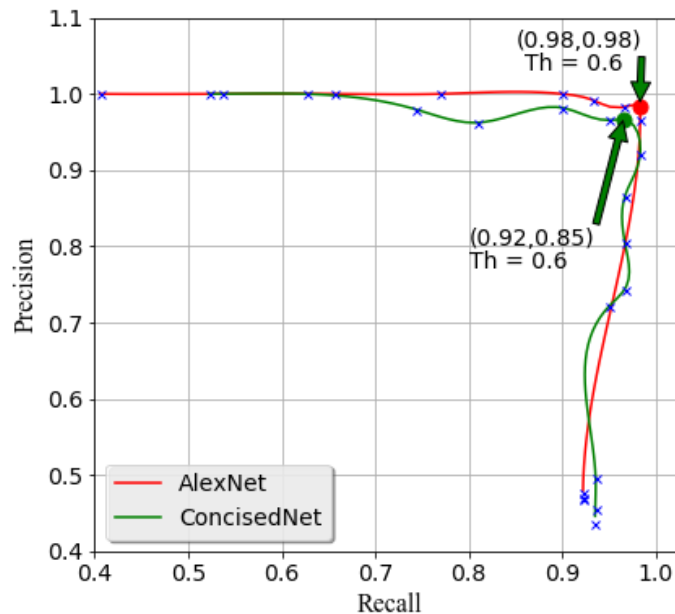


Figure 6. Using AlexNet and ConcisedNet, Precision vs Recall for different values of the threshold for Talking Face dataset are shown. F1 score shows a Threshold = 0.5 and 0.6 can provide the best results for AlexNET and ConcisedNet, respectively.

The performance of the proposed method is listed as below in table 4 and table 5. Here, PR: Precision, RE: Recall, GT: Ground Truth, TP: True Positive, FP: False Positive. FN: False Negative. The obtained results show high detection accuracy of the blink action for the two tested networks: the ConcisedNet-RNN and the AlexNet-RNN. In fact, the two networks provided very similar results on the Talking Face dataset of about 97-98% in both precision and recall. However, AlexNet-RNN pipeline provides better result on the Eyeblink8 when compared to ConcisedNet-RNN. As it has been already mentioned, the implementation of ConcisedNet was tested mainly to address the real-time processing requirements with the software running on a CPU, while AlexNet need a GPU to work in the real-time regime.

Table 4: Shows the results obtained using ConcisedNet with the RNN for the two datasets, EyeBlink8 and Talking Face.

| Dataset | PR % | RE % | GT | TP | FP | FN |
|-----------|-------|------|-----|-----|----|----|
| TF | 97 | 97 | 59 | 57 | 2 | 2 |
| Eyeblink8 | 75.35 | 84 | 255 | 214 | 70 | 41 |

Table 5: Shows the results obtained for AlexNET. The RNN model provided better results with AlexNET than the ConcisedNet regarding the two datasets.

| Dataset | PR % | RE% | GT | TP | FP | FN |
|-----------|------|------|-----|-----|----|----|
| TF | 98.3 | 98.3 | 59 | 58 | 1 | 1 |
| Eyeblink8 | 93.6 | 93.9 | 255 | 241 | 17 | 14 |

6. CONCLUSION

This paper reports on the work, investigating software solution that can be used for spatio-temporal motion detection and quantification from video. The solution consists of a CNN followed by a multilayer RNN. To prove the concept, temporal sequences obtained from the classification stage of the eye states detection are used to train multilayer RNN to predict eye blinks in videos. The CNN is trained for eye status detection using the Helen Dataset. The CNN generated classification descriptors are sequenced as an input to the RNN to provide a prediction of an eyelid motion and subsequently used to detect eye-blink events. To show that the RNN network is not too dependent on a specific CNN architecture that provide the training sequence, only the ConcisedNet is used to provide the training sequence. Then during the inference stage, both AlexNet and ConcisedNet are used in two separate experiments to obtain the performance measures. It had been shown that when the RNN is attached to the AlexNET it provides better results. On the Talking Face dataset, the proposed solution achieved 98.3% for both precision and recall, while for the Eyeblink8 dataset the obtained result were 93.6% and 93.9% for precision and recall respectively.

The main advantage of this approach is that it does not require prior knowledge about the testing environment. If the CNN can provide the necessary detection sequences, it is expected that the RNN would perform the eye blink detection.

REFERENCES

- [1] Anas R. E., Henriquez P. and Matuszewski, J. B. (2017). Online Eye Status Detection in the Wild with Convolutional Neural Networks. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6: VISAPP, (VISIGRAPP 2017)* ISBN 978-989-758-227-1, pages 88-95.
- [2] Karpathy, A. and Fei-Fei, L., 2015. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3128-3137).
- [3] Drutarovsky, T., Fogelton, A., 2014. Eye blink detection using variance of motion vectors. In: *Computer Vision - ECCV 2014 Workshops*, Zurich, Switzerland. Springer, Cham, Switzerland, pp. 436-448.
- [4] Fogelton, A., Benesova W., 2016. Eye blink detection based on motion vectors analysis. *Computer Vision and Image Understanding*, 148, 23-33.
- [5] Portello, JK., Resenfield, M., Chu, CA., 2013. Blink Rate, Incomplete Blinks and Computer Vision Syndrome. *Optometry and Vision Science*, 90 (5), 4 82-4 87.
- [6] Du, Y., Ma, P., Su, X., Zhang, Y., 2008. Driver Fatigue Detection based on Eye State Analysis. In: *Proceedings of the 11th Joint Conference on Information Sciences*.
- [7] Mohammed, A., Anwer, S., 2014. Efficient Eye Blink Detection Method for disabled-helping domain. *Inter. Jour. of Adv. In: Computer Science and Applications*.
- [8] Joshi, KV., Kangda, A., Patel, S., 2016. Real Time System for Student Fatigue Detection during Online Learning", *Inter. Jour of Hybrid Info. Tech.*
- [9] Rezaei, M., Klette, R., 2012. Novel Adaptive Eye Detection and Tracking for Challenging Lighting Conditions. *Asian Conference on Computer Vision*, 427-440.
- [10] Sun, Y., Zafeiriou, S., Pantic, M., 2013. A Hybrid System for On-line Blink Detection. In: *Hawaii International Conference on System Sciences*.
- [11] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [12] Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T., 2012. Interactive facial feature localization. *ECCV'12*, pp.679-692.

Study on Feature Extraction of Bearing Incipient Weak Fault Based on Dictionary Learning

Guangming Dong¹, Jin Chen¹, Haodong Yuan¹

¹ State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, No.800, Dongchuan Rd, Shanghai 200240, China

ABSTRACT

Rolling element bearings are frequently used in rotary machinery, but they are also fragile mechanical parts. Hence, exact condition monitoring and fault diagnosis for them plays an important role of ensuring machinery's reliable running. Timely diagnosis of early bearing faults is desirable, but the early fault signatures are easily submerged in noise. This paper focuses on the application of dictionary learning and sparse representation methods on bearing fault feature extraction and fault diagnosis. Two dictionary learning methods: K-SVD and shift-invariant dictionary learning (SIDL) are studied. The K-SVD algorithm takes its name from singular value decomposition (SVD), which is used for one atom update and repeated for K times during the dictionary learning stage. K-SVD is one appealing method because of its efficiency. SIDL reconstructs signals using basis atoms in all possible time shifts. SIDL is very suitable to extract periodic impulses. Simulation and experimental bearing signals are used for demonstration and comparison of fault feature extraction based on K-SVD and SIDL.

Keywords: Feature extraction; sparse representation; Dictionary learning; Shift-invariant dictionary learning; K-SVD.

Corresponding author: Guangming Dong (gmdong@sjtu.edu.cn)

1. INTRODUCTION

Rolling element bearings have a wide range of applications in rotating machinery, while they are easily damaged, causing serious mechanical failure. Bearing fault diagnosis is an important issue in the industrial manufacturing process. Vibration monitoring is a commonly used way to detect the bearing conditions. For machinery fault diagnosis based on vibration signals, there are always two main steps, i.e., feature extraction and pattern recognition. With respect to feature extraction of mechanical vibration signals, researchers have proposed many methods based on the time domain, frequency domain or time-frequency analysis, such as spectral kurtosis [1], cyclostationary analysis [2], wavelet transform [3], empirical mode decomposition [4], etc.

Sparse representation is a relatively new signal processing method[5]. Compared with the methods mentioned above, sparse decomposition provides much more succinct representations aiming to seek sparsest or nearly-sparsest representations of signals over particular over-complete dictionaries. The choice of the overcomplete dictionary is an important issue for sparse representation of signals[6]. Study on this topic can be divided into two categories: the analytic approach and the learning-based approach. The first approach chooses a pre-defined dictionary, such as Fourier dictionary, wavelet dictionary and Gabor dictionary. This kind of approaches has relatively simple algorithms because of the rigorous dictionary structure. The sparse representation methods based on pre-defined dictionaries have been introduced for machinery fault diagnosis[7, 8]. The second approach is to find the over-complete dictionary that can sparsely represent the training signals by machine learning techniques. Unlike the sparse representations based on predefined analytic dictionaries, the approaches based on dictionary learning allow more flexibility to adapt to variation of real signals. The research in dictionary learning has three main categories[9]: i) the probabilistic learning methods, such as methods using maximum likelihood[10]; ii) the learning methods based on clustering or vector quantization, such as MOD (method of optimal

directions)[11] and K-SVD(K-singular value decomposition)[12]; iii) the methods for learning dictionaries with a particular construction, such as shift invariant representation or SIDL[13].

In this paper, K-SVD and SIDL is used for the sparse representation of bearing signals with incipient faults. The K-SVD algorithm takes its name from SVD, which is used for one atom update and repeated for K times during the dictionary learning stage. K-SVD is one appealing method because of its efficiency. SIDL reconstructs signals using basis atoms in all possible time shifts. SIDL is very suitable to extract periodic impulses. Simulation and experimental bearing signals are used for demonstration and comparison of fault feature extraction based on K-SVD and SIDL.

2. SPARSE REPRESENTATION, K-SVD AND SIDL

2.1. Sparse Representation of Signals

The goal of sparse representation is to construct a signal as a linear combination of a small number of atoms from a dictionary. Elements of the dictionary are typically unit norm functions called atoms.

Let us denote an overcomplete dictionary matrix $\mathbf{D} \in \mathbb{R}^{n \times K}$ ($n < K$) that contains K atoms as \mathbf{D} 's columns $\{\mathbf{d}_j\}_{j=1, \dots, K} \in \mathbb{R}^{n \times 1}$. A signal $\mathbf{y} \in \mathbb{R}^{n \times 1}$ can be represented as a linear combination of these atoms. The representation of \mathbf{y} is approximated as $\mathbf{y} \approx \mathbf{D}\boldsymbol{\alpha}$, with approximation error $\|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2 < \varepsilon$. The vector $\boldsymbol{\alpha} \in \mathbb{R}^{K \times 1}$ contains the representation coefficients of the signal \mathbf{y} . In the approximate representation, $\|\cdot\|_2$ stands for a l_2 norm of the error vector $\|\mathbf{e}\|_2 = \sqrt{\sum_i e_i^2}$. If the overcomplete dictionary \mathbf{D} is a full rank matrix, an infinite number of solutions for $\boldsymbol{\alpha}$ are available for the representation problem. The sparsest representation problem can be stated as follows:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{subject to } \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2 < \varepsilon \quad (21)$$

Where $\|\cdot\|_0$ is the l_0 norm, counting the nonzero elements in a vector. Solution of the sparsest representation for equation (21) has been extensively investigated in the past few years. The choice of the overcomplete dictionary \mathbf{D} is an important issue for sparse representation of signals[6]. Study on this topic can be divided into two categories: the analytic approach and the learning-based approach. In the first approach, a mathematical model of the dictionary is pre-defined, such as Fourier dictionary, wavelet dictionary and Gabor dictionary. The success of such dictionaries in applications depends on how suitable they are to sparsely describe the target signals. The second approach is to find the dictionary \mathbf{D} that can sparsely represent the training signals by machine learning techniques. Unlike the sparse representations based on predefined analytic dictionaries, the approaches based on dictionary learning allow more flexibility to adapt to variation of real signals.

2.2. Dictionary learning theory and K-SVD

By using techniques of signal processing and machine learning, dictionary learning aims at finding a dictionary \mathbf{D} from the training data set for a sparse representation of target signals. Let the training signal set be $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M] \in \mathbb{R}^{n \times M}$, formed by M one-dimensional signals as its columns. Let $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M] \in \mathbb{R}^{K \times M}$ be a matrix containing the sparse coefficient vectors $\boldsymbol{\alpha}_j$ of all the input signals \mathbf{y}_j in \mathbf{Y} . A commonly used formulation for dictionary learning is the following optimization problem:

$$\min_{\mathbf{D}, \mathbf{A}} \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 \quad \text{subject to } \|\boldsymbol{\alpha}_j\|_0 < k_0, \quad 1 \leq j \leq M \quad (22)$$

Where $\|\mathbf{M}\|_F$ stands for the Frobenius norm of matrix, $\|\mathbf{M}\|_F = \sqrt{\sum_i \sum_j m_{ij}^2}$.

The process of solving the dictionary learning problem of equation (22) is an iterative alternation between the sparse coding stage and the dictionary update stage, as shown in figure 1.

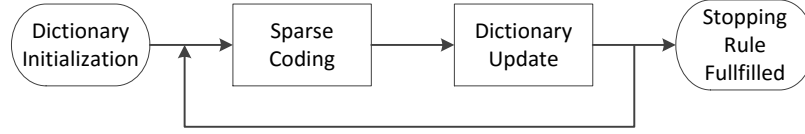


Figure 1. The process of dictionary learning

As shown in figure 1, after initialization of the dictionary matrix \mathbf{D} , the first step in dictionary learning is the sparse coding stage, where \mathbf{D} is known and fixed. In this stage, the optimization problem of equation (22) can be decoupled to M distinct problems of the form:

$$\min_{\alpha_j} \|\mathbf{y}_j - \mathbf{D}\alpha_j\|_2^2 \quad \text{subject to } \|\alpha_j\|_0 < k_0, \quad 1 \leq j \leq M \quad (23)$$

This problem can be solved by the pursuit algorithms, such as Matching Pursuit (MP)[14], Orthogonal Matching Pursuit (OMP) or Basis Pursuit(BP) [15] by replacing the l_0 norm in equation (23) with an l_1 norm.

The second step in dictionary learning is the dictionary update stage, where the coefficient matrix \mathbf{A} is fixed as known. In this stage, the optimization problem of equation (22) becomes

$$\mathbf{D} = \arg \min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 \quad (24)$$

Taking the derivative of $\|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2$ in equation (24) with respect to \mathbf{D} , we obtain $(\mathbf{Y} - \mathbf{D}\mathbf{A})\mathbf{A}^T$, leading to

$$\mathbf{D} = \mathbf{Y}\mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{Y}\mathbf{A}^+ \quad (25)$$

The above equation (25) need to compute the pseudo inverse of the matrix \mathbf{A} , which is time consuming because \mathbf{A} is often very large. The K-SVD algorithm proposed in [12], avoids the matrix inverse computation in equation (25) by updating the dictionary \mathbf{D} one atom at a time. In K-SVD, the multiplication $\mathbf{D}\mathbf{A}$ in equation (24) is decomposed to the sum of K rank-1 matrices as:

$$\mathbf{D}\mathbf{A} = \sum_{j=1}^K \mathbf{d}_j \alpha_T^j = \sum_{j \neq k} \mathbf{d}_j \alpha_T^j + \mathbf{d}_k \alpha_T^k \quad (26)$$

In equation (26), \mathbf{d}_j is the j -th column or atom in the dictionary \mathbf{D} , and α_T^j is the j -th row in the coefficient matrix \mathbf{A} . During the dictionary upstate stage in K-SVD, $K-1$ atoms in \mathbf{D} are assumed fixed; while one atom, the k -th, \mathbf{d}_k is to be updated.

The representation error $\|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2$ can be written as:

$$\|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F^2 = \|\mathbf{E}_k - \mathbf{d}_k \alpha_T^k\|_F^2 \quad (27)$$

In equation (27), the matrix \mathbf{E}_k stands for the error for the training signal set when the k -th atom \mathbf{d}_k is removed. The singular value decomposition (SVD) can be used to find the closest rank-1 matrix that approximates \mathbf{E}_k , through which the removed \mathbf{d}_k can be calculated. However, direct SVD of \mathbf{E}_k will affect

the sparsity structure of α_T^k while updating d_k . To enforce the sparsity constraint on α_T^k , define ω_k as the set of indices for nonzero $\alpha_T^k(i)$:

$$\omega_k = \{i | 1 \leq i \leq K, \alpha_T^k(i) \neq 0\} \quad (28)$$

Define Ω_k as a matrix of size $M \times \|\omega_k\|_0$, with ones on the $\{\omega_k(i), i\}$ entries, and zeros elsewhere. Using Ω_k as a sparsity constraint:

$$\|\mathbf{E}_k \Omega_k - d_k \alpha_T^k \Omega_k\|_F^2 = \|\mathbf{E}_k^R - d_k \alpha_R^k\|_F^2 \quad (29)$$

In equation (29), $\alpha_R^k = \alpha_T^k \Omega_k$ shrinks the row vector α_T^k by discarding the zero elements. For the restricted error matrix \mathbf{E}_k^R , SVD can be used:

$$\mathbf{E}_k^R = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad (30)$$

From equation (30), we can obtain the k-th updated atom \tilde{d}_k from the first column of \mathbf{U} , and obtain the updated coefficient $\tilde{\alpha}_R^k$ from the first column of \mathbf{V} multiplied by $\mathbf{\Lambda}(1,1)$:

$$\begin{aligned} \tilde{d}_k &= \mathbf{U}(:,1) \\ \tilde{\alpha}_R^k &= \mathbf{V}(:,1) \mathbf{\Lambda}(1,1) \end{aligned} \quad (31)$$

2.3. Shift-Invariant Dictionary Learning

In shift-invariant dictionary learning (SIDL), each basis function is allowed to appear at all possible shifts within the signal. Our goal is to model a signal by using a small set of kernel functions that can be placed at arbitrary time points. Ultimately, we want to find the minimal set of functions and time points that fit the signal within a given noise level. We expect this type of model to work well for signals composed of events whose onset can occur at arbitrary temporal positions. We assume time series $y(t)$ is modelled by

$$y(t) = \sum_i a_i d_{m[i]}(t - \tau_i) + \varepsilon(t) \quad (32)$$

Where τ_i indicates the temporal position of the i^{th} basis function $d_{m[i]}$, which is scaled by a_i . The notation $m[i]$ represents an index function that specifies which of the M kernel functions is present at time τ_i . A single basis function can occur at multiple times during the time series. Additive noise at time t is given by $\varepsilon(t)$. A more general way to express (32) is to assume that the basis functions exist at all time points during the signal, and let the non-zero coefficients determine the positions of the basis functions. In this case, the model can be expressed in convolutional form, where $a_m(t)$ is the coefficient at time t for basis function d_m .

$$y(t) = \sum_m \int a_m(\tau) d_m(t - \tau) d\tau + \varepsilon(t) = \sum_m d_m(t) * a_m(t) + \varepsilon(t) \quad (33)$$

The corresponding SIDL optimization problem for dictionary learning and coefficients updating then becomes:

$$\min_{d,a} \sum_{i=1}^m \left\| x^{(i)} - \sum_{j=1}^n d^{(j)} * a^{(i,j)} \right\|_2^2 + \beta \left\| \sum_{i,j} a^{(i,j)} \right\|_1 \quad \text{s.t.} \quad \|d^{(j)}\|_2^2 \leq c, \quad 1 \leq j \leq n \quad (34)$$

We solve the optimization problem (34) by alternately solving two large convex optimization problems: computing the coefficients $a^{(i,j)}$ given a fixed dictionary, and finding the bases $d^{(j)}$ given the coefficients.

Given a fixed set of dictionary, the coefficients $\{s^{(i,j)}, j=1, \dots, n\}$ corresponding to an input $x^{(i)}$ can be found independently of the coefficients corresponding to other inputs. Thus, we will focus on the problem of finding the coefficients a corresponding to a single input $x \in R^p$:

$$\min_a \left\| x - \sum_{j=1}^n d^{(j)} * a^{(j)} \right\|_2^2 + \beta \left\| \sum_{i,j} a^{(j)} \right\|_1 \quad (35)$$

Here, $a^{(j)} \in R^{p-q+1}$ represents the coefficients associated with $d^{(j)}$. Optimizing (35) with respect to $a^{(j)}$ is difficult because the variables are highly coupled. Notice now that if we know the signs (positive, zero, or negative) of the $a^{(i,j)}$ at the optimal value, we can replace each of the terms $|a^{(i,j)}|$ with either $a^{(i,j)}$ (if $a^{(i,j)} > 0$), $-a^{(i,j)}$ (if $a^{(i,j)} < 0$), or 0 (if $a^{(i,j)} = 0$). Our approach will efficiently find the exact optimum with respect to all the coefficients by taking advantage of the “feature-sign” trick. This feature-sign trick has been formalized in the feature-sign search algorithm, which greedily searches the space of nonzero coefficients and their signs, and provably converges to the optimal solution.

Solving the SIDL optimization problem (35) for the basis $d^{(j)}$ keeping the coefficients a fixed reduces to an L_2 -constrained optimization problem:

$$\min_a \left\| x - \sum_{j=1}^n d^{(j)} * a^{(j)} \right\|_2^2 \quad \text{s.t.} \quad \|d^{(j)}\|_2^2 \leq c, \quad 1 \leq j \leq n \quad (36)$$

In this problem, because each basis function can appear in any possible shift, each component of the basis vector contributes to many different terms in the objective function. Therefore, unlike regular sparse coding, this problem is much easier to solve when transformed into the frequency domain. Let the discrete Fourier transform (DFT) be denoted by $\hat{d}^{(j)}$, ($j=1, 2, \dots, p$), the optimization problem (36) becomes:

$$\min_{\hat{d}^{(j)}} \left\| \hat{x}^{(i)} - \sum_j \hat{d}^{(j)} \cdot \hat{a}^{(i,j)} \right\|_2^2 \quad \text{s.t.} \quad \|\hat{d}^{(j)}\|_2^2 \leq \hat{c} = cK \quad (37)$$

where $\hat{x}^{(i)} \in \mathbb{C}^p$ and $\hat{a}^{(i,j)} \in \mathbb{C}$ represent the complex-valued DFT for the input $x^{(i)}$ and coefficients $a^{(i,j)}$ respectively. The problem of (37) can be solved using method described in Ref.[13].

3. SIMULATED SIGNAL ANALYSIS OF FAULTED BEARINGS BASED ON K-SVD AND SIDL

3.1. Simulation model of faulted bearings

The vibration signal of rolling element bearing includes components resulting from its structure, tolerance, and surface deterioration such as spalling fatigue and abrasive wear. McFadden[16] developed a model produced by a single point defect, which incorporates the effects of rolling element bearing geometry, shaft speed, load distribution, transfer function, and the exponential decay vibration. This model was extended to describe the vibration produced by multiple point defects [17]. The models above consider shocks as exactly periodicity. However, because the rolling elements experience some random slip, these shocks' periodicity is not exact. In addition, their amplitudes are subjected to modulation by the rotations of the inner race (inner race fault), the outer race (outer race fault), or the cage (rolling element fault). Synthetically considering these factors, the mathematical model of defective rolling element bearing can be expressed as:

$$\begin{aligned}
 x(t) &= \sum_{i=1}^M A_i \cdot s(t - iT - \tau_i) + n(t) \\
 A_i &= A_0 \cdot \cos(2\pi f_m t + \varphi_A) + C_A \\
 s(t) &= e^{-Bt} \cdot \cos(2\pi f_n t + \varphi_w)
 \end{aligned} \tag{38}$$

where A_i is the amplitude modulator with period $Q=1/f_m$, $s(t)$ the attenuated damping oscillation with mean impulse period T , τ_i the minor fluctuation around T , $n(t)$ the Gaussian noise with zero-mean, f_n the natural frequency related to bearing or system, A_0 the resonance intensity, C_A the arbitrary constant, B is the coefficient of resonance damping, depending on system, φ_A and φ_w are the initial phases of the amplitude modulator A_i and the attenuated damping oscillation $s(t)$.

3.2. Bearing fault feature extraction based on K-SVD and SIDL

In simulation, the data length of $x(t)$ is 25600, the sampling frequency is 25.6 kHz. At first, the signal-to-noise ratio is set as SNR=-2dB. In K-SVD analysis, $x(t)$ is segmented to obtain the training signal set, where the data segment length is 128, the number of overlap between the adjacent data segment is 32, and the number of atoms in dictionary is set as K=160. After training, time waveforms of the randomly selected nine atoms are shown in figure 2, from which it can be seen that atoms of No. (1)(2)(5)(8)(9) have obvious impulsive patterns.

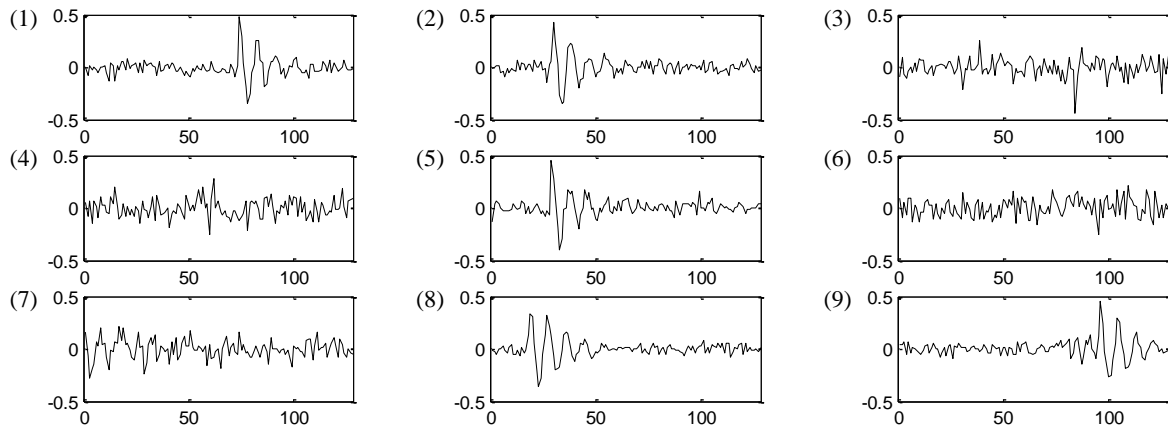


Figure 2. Nine randomly selected atoms from the learned dictionary

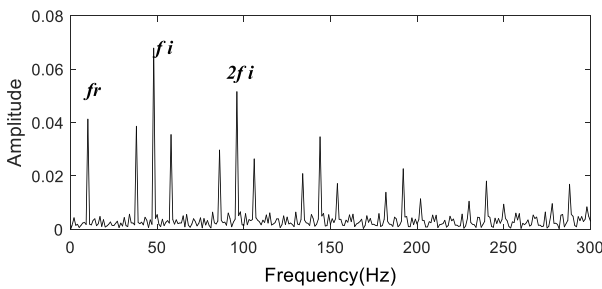


Figure 3. Envelope spectrum of the de-noised signal using K-SVD (SNR = -2dB)

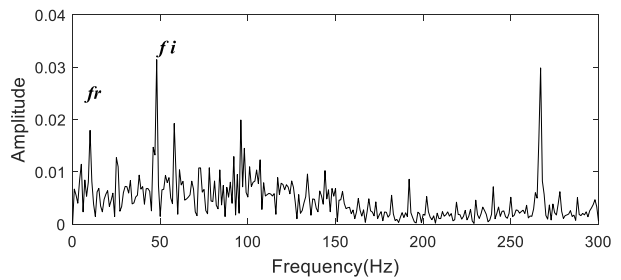


Figure 4. Envelope spectrum of the de-noised signal using K-SVD (SNR = -10dB)

By using the learned dictionary, figure 3 shows the envelope spectrum of the de-noised signal, when SNR is -2dB. It can be seen from figure 3 that the bearing fault feature can be clearly extracted. When the noise level is increased, figure 4 shows the result when SNR is decreased to -10dB. The noise influences is evident when using K-SVD.

The simulated signal with SNR -10dB is analyzed using SIDL. The length of dictionary atom for training is set as 256, which is long enough to describe single feature in this case, such as one impulse. The number of atom in the dictionary is set as 4. Figure 5 shows the four atoms learned from simulated signal (SNR -10dB). It can be seen from figure 5 (a) that the exponentially decaying impulse structure can be obviously observed, which is exact the signal structure of the simulated bearing fault. Signal is reconstructed using the atom of figure 5 (a) with the method of SIDL, the envelope spectrum of the reconstructed signal is shown in figure 6. Comparing figure 6 and figure 4, it can be seen that the SIDL is more powerful than K-SVD in extracting bearing fault under heavy noise. However, SIDL is only suitable for extracting fault feature with repeated features.

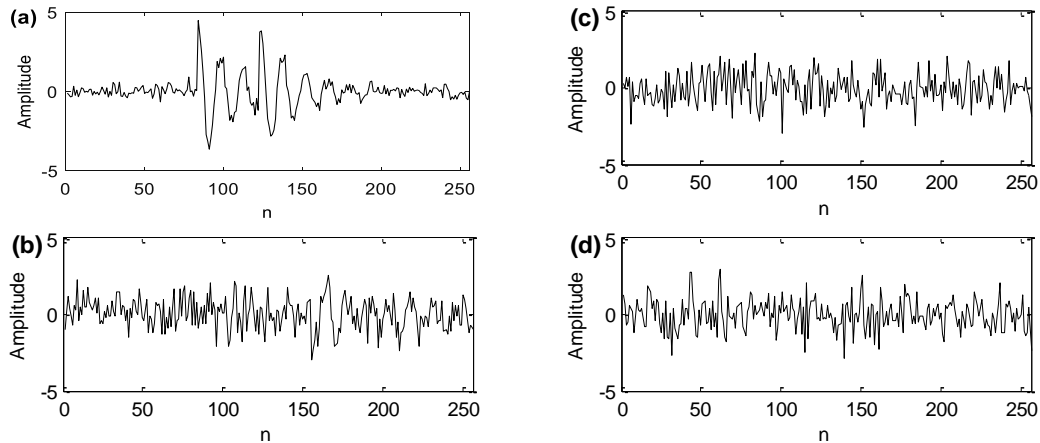


Figure 5. The four learned atoms of bearing fault signal using SIDL

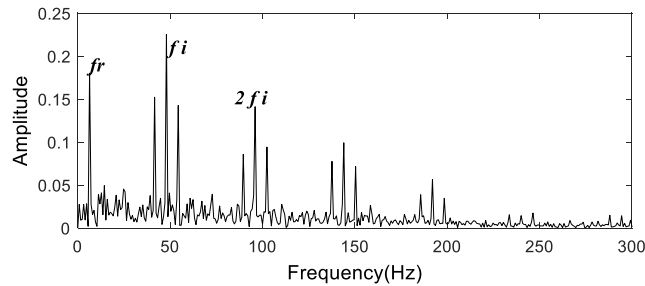


Figure 6. Envelope spectrum of the reconstructed signal by atom (a) using SIDL (SNR = -10dB)

4. EXPERIMENTAL STUDY

The rolling element bearing accelerated life test is carried out on accelerated bearing life tester (ABLT-1A) which simultaneously hosts four rolling element bearings on one shaft driven by an AC motor and coupled by rubber belts. 20480 points data of acceleration signal is collected per minute at 25.6KHz sampling rate. The rotational speed of the shaft is 3000 rpm.

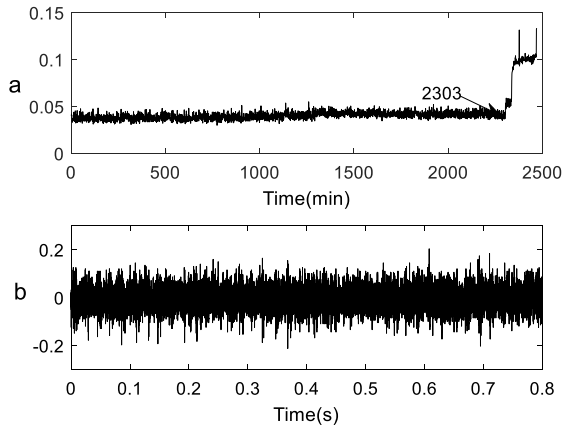


Figure 7. (a) RMS over whole life time, (b) raw signal at 2303 minute

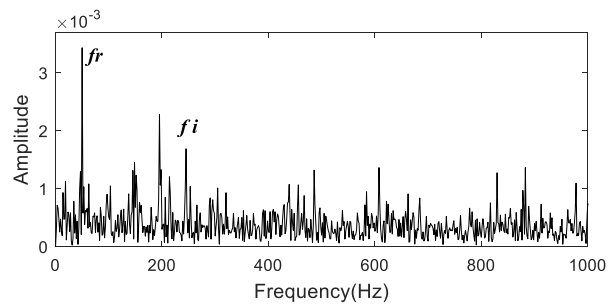


Figure 8. Envelope spectrum of the raw signal at 2303 minute

Figure 7(a) shows the general trend of RMS over whole life time. After 2303 min, it shows an obvious jump which can be seen as a start point of apparent defect. In order to realize predictive maintenance and reduce the occurrence of abrupt failure, we choose the raw signal collected at 2303 min which is displayed in figure 7(b) to realize the early fault feature extraction. Envelope spectrum of the raw signal at 2303 minute is shown in figure 8, where the noise effect is obvious. The experimental signal is processed using K-SVD and SIDL respectively. The envelope spectra of the processed signals using two methods are shown in figure 9 and 10. In K-SVD, atom length is chosen as 64 for a steady result, while atom length is chosen as 256 for an optimal result using SIDL. It can be seen that SIDL obtains a better results, while K-SVD is questionable for feature interpretation using figure 9.

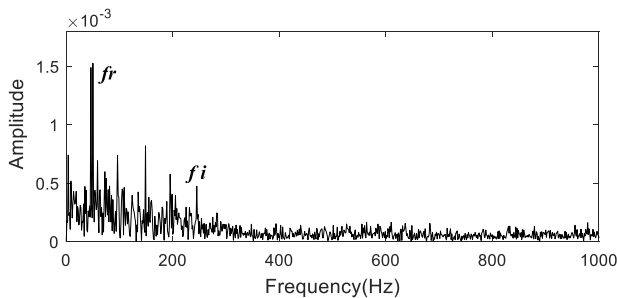


Figure 9. Envelope spectrum of the denoised signal using K-SVD

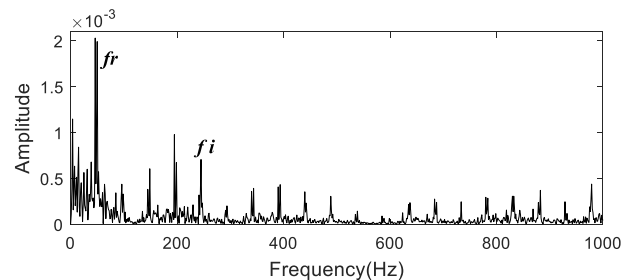


Figure 10. Envelope spectrum of the reconstructed signal using SIDL

5. CONCLUSION

In this paper, two methods based on dictionary learning and sparse representation are studied, which are K-SVD and shift-invariant dictionary learning (SIDL). Their applications in fault feature extraction of rolling element bearings are demonstrated by simulation and experimental signals. Although K-SVD is more popular in sparse representation community, the fault feature extraction ability of SIDL is more powerful than K-SVD for bearing signals under heavy noise.

REFERENCES

- [1] Antoni, J., Randall, R. B. (2006) The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*. 20(2): 308-331.
- [2] Randall, R. B., Antoni, J., Chobsaard, S. (2001) The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mechanical Systems and Signal Processing*. 15(5): 945-962.

- [3] Feng, Y. H., Schlindwein, F. S. (2009) Normalized wavelet packets quantifiers for condition monitoring. *Mechanical Systems and Signal Processing*. 23(3): 712-723.
- [4] Yu, D. J., Cheng, J. S., Yang, Y. (2005) Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical Systems and Signal Processing*. 19(2): 259-270.
- [5] Olshausen, B. A. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 381(6583): 607-609.
- [6] Rubinstein, R., Bruckstein, A. M., Elad, M. (2010) Dictionaries for Sparse Representation Modeling. *Proceedings of the Ieee*. 98(6): 1045-1057.
- [7] Feng, Z. P., Chu, F. L. (2007) Application of atomic decomposition to gear damage detection. *Journal of Sound and Vibration*. 302(1): 138-151.
- [8] Cui, L., Wang, J., Lee, S. (2014) Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis. *Journal of Sound and Vibration*. 333(10): 2840-2862.
- [9] Tomic, I., Frossard, P. (2011) Dictionary Learning. *Ieee Signal Processing Magazine*. 28(2): 27-38.
- [10] Olshausen, B. A., Field, D. J. (1997) Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*. 37(23): 3311-3325.
- [11] Engan, K., Aase, S. O., Husoy, J. H., Method of optimal directions for frame design, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, 2443-2446.
- [12] Aharon, M., Elad, M., Bruckstein, A. (2006) K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Signal Processing, IEEE Transactions on*. 54(11): 4311-4322.
- [13] Blumensath, T., Davies, M. (2006) Sparse and shift-invariant representations of music. *IEEE Transactions on Audio Speech and Language Processing*. 14(1): 50-57.
- [14] Mallat, S. G., Zhang, Z. (1993) Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*. 41(12): 3397-3415.
- [15] Chen, S. S. B., Donoho, D. L., Saunders, M. A. (1998) Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*. 20(1): 33-61.
- [16] McFadden, P. D., Smith, J. D. (1984) Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*. 96(1): 69-82.
- [17] McFadden, P. D., Smith, J. D. (1985) The vibration produced by multiple point defects in a rolling element bearing. *Journal of Sound and Vibration*. 98(2): 263-273.

Session 16

Monitoring, Diagnosis, Prognosis and Health Management 4

Session Chair Dr. Violet Leavers

The Impact of Effective Prognostic Techniques (Predicting Remaining Useful Life) on Successful Implementation of Total Productive Maintenance in the Power Industry

(A. Alseiri, P. Farrell)

Fault Diagnosis of a Polymer Electrolyte Membrane Fuel Cell Using Bayesian Network

(L. Mao, L. Jackson, B. Davies)

Weighted Narrowband Envelope Spectrum and its Application to Bearing Fault Diagnosis

(J. Duan, T. Shi, J. Xuan, H. Zhou)

The Impact of Effective Prognostic Techniques (Predicting Remaining Useful Life) on Successful Implementation of Total Productive Maintenance in the Power Industry

A.Y. Alseiyari¹, P. Farrell²

¹ Planning and Development Directorate, Abu Dhabi Water and Electricity Authority, Abu Dhabi 6120, UAE

² School of Engineering, University of Bolton, Bolton BL3 5AB, UK

ABSTRACT

The power industry has recognised the significance of implementing Total Productive Maintenance (TPM) as a maintenance strategy to enhance its proactive techniques and improve power asset reliability. Successful power utilities have sufficient failure data that predicts accurate times to failure of their power assets, based on condition monitoring techniques. Power transformers are the most expensive elements and their faults are particularly costly. Therefore, detection of power transformer failures at an early stage is important to assure reliable and efficient operation. The aim of this research is to investigate the critical role of prognostic techniques and how they can drive successful implementation of TPM in the power industry through planned maintenance pillar. The research presents an innovative technique for early fault detection and diagnosis in power transformers based on paper degradation analysis. This method uses measurements obtained from a power distribution industry in the Abu Dhabi of the degradation factor of power transformer paper insulation (degree of polymerisation - DP values) over time, from condition monitoring data that is obtained through dissolved-gas analysis (DGA). The mathematical exponential model uses degradation measurements DP values of paper insulation for multiple power transformers to extrapolate these measurements over time, and hence predicts potential failure times before they occur. This is based on a prognostic technique that utilises ReliaSoft's Weibull++ software. The results of this research show that successful implementation of TPM in the power industry can be achieved through appropriate techniques of predictive maintenance within the planned maintenance pillar of TPM, thus increasing availability, performance, quality and reliability rates by preventing sudden failure of power transformers.

Keywords: Total Productive Maintenance (TPM); Condition Monitoring (CM); Predictive Maintenance (PdM); Dissolved-Gas Analysis (DGA); Degradation Analysis; Remaining Useful Life (RUL)

Corresponding author: A. Alseiyari (email: alsayaaari@hotmail.com)

1. INTRODUCTION

Nowadays, maintenance has a significant impact on the business strategy of power organisations. Maintenance of power plants aims to promote safe, cost effective, and reliable electricity supply. This kind of maintenance has complex activities and requires highly skilled maintenance staff; therefore, proper utilisation of technology is critical to achieve both plant safety and availability [1]. Power companies have recognised the importance of implementing Total Productive Maintenance (TPM) as a maintenance strategy to optimise their proactive techniques and improve power asset reliability. Further, successful power utilities need to have robust data management procedures such as computerised maintenance management systems (CMMSs), that includes sufficient failure data. Otherwise, these utilities cannot predict accurate times to failure of their power assets through condition monitoring (CM) techniques. CM systems lead the power industry to better maintenance management performance and higher reliability. CM can monitor the critical parts of power plants such as transformers, and determine the optimal period for particular maintenance [2]. Preventive maintenance involves the systematic scheduling of inspection, detection, and prevention of incipient faults before they occur as major or actual failures. Predictive maintenance (PdM) is a distinctive type of preventive maintenance that can be performed at intervals or continuously, based on

the requirements to diagnose the condition of equipment. PdM plays an essential role to generate improvement in safety, availability, performance, quality of equipment and reduced maintenance costs. Implementation of PdM becomes significant where security of power plants is critical. Thus, it has been utilised as a strategic decision making technique that consists of predictive tools as well cost-effective methods (e.g. dissolved-gas analysis, vibration monitoring, and thermography) to obtain real time data and hence measure the physical condition of equipment during its operation. According to this, all activities of maintenance are organised and scheduled, based on priority and criticality analysis of equipment [3]. The purpose of this paper is to investigate the critical role of prognostic techniques and how they can drive successful implementation of TPM in the power industry through planned maintenance pillar. With high demand on quality, productivity, and availability, power plants become more complex and costly. Developing a maintenance programme is not an easy process, and it faces many obstacles during its implementation; it often lacks a systematic and reliable methodology. The finding in this paper is that there is a need to apply effective prognostic techniques to drive successful implementation of TPM in the power industry through PdM programmes.

2. LITERATURE REVIEW

2.1. Total Productive Maintenance (TPM)

TPM is an innovative maintenance strategy employed to increase overall equipment effectiveness by establishing an inclusive productive maintenance scheme covering the entire life-cycle of the equipment [4]. It brings operators and maintenance staff together through creating a combination of good working procedures with high standards, team working, and focused improvement, thus increasing the effectiveness of processes and equipment [5] – [6]. It was introduced in Japan 1971 by Nakajima [4] who was the chairman of the Japanese Institute of Plant Maintenance (JIPM) and is extensively adopted in manufacturing sectors today. TPM aimed to achieve zero equipment failures, zero accidents, and zero defects in order to have high availability, performance, and quality. The principle TPM practices are named the pillars of TPM, specifically: planned maintenance; focused improvement; autonomous maintenance; development management; office TPM; education and training; quality maintenance; and safety, health and environment. An explanation about TPM process can be found in Nakajima [4] - Reiman and Oedewald [1]. Operators and maintenance staff need to have high levels of knowledge, be able to share each other's job and sometimes develop new skills to achieve the TPM philosophy [6]. Its implementation is not easy and faces many obstacles; details about its barriers are articulated by Ahuja and Khamba [7]. Waeyenbergh and Pintelon [8] argued that TPM does not reflect a maintenance concept because it has no clear maintenance rules and policies to decide which simple maintenance policy will be utilised. Thus, building a robust planned maintenance pillar in the power industry needs to be investigated to promote TPM implementation through predictive maintenance tools.

2.2. Predictive maintenance (PdM)

Predictive maintenance techniques has become the foundation for most modern maintenance methods. Predictive maintenance programs involve different methods of problem diagnosis in order to predict equipment faults before they occur. This diagnosis involves several types of equipment testing to deliver results that can be compared with standard results, and based on results, correct maintenance action can be taken [9]. Condition monitoring techniques are: vibration analysis, acoustic analysis, thermography-infrared radiation, oil analysis and human senses among others. Further, the Dissolved-Gas Analysis (DGA) technique examines the oil and its chemical composition in order to predict failure modes. The results of this test will determine the following: high content of silicon thus indicating contamination of grit; and high levels of iron indicating a wearing of components, and iron or aluminium presence indicating worn piston rings. Right fault classification is significant in power transformers to minimise power quality (PQ) disturbances. Partial discharge (PD), corona, arcing and thermal heating cause PQ disturbances and hence may lead to catastrophic failures (e.g. interruption of power supply, damaged equipment, increase maintenance costs, and high safety risks to operators and technicians). Gases generate when faults get dissolved in the oil of power transformers. DGA is widely utilised by power utilities as an effective tool for

fault diagnosis of transformers. Since the late 1970s, DGA interpretation has used standards such as IEC 60599, IEEE C57.104-1991, etc [10]. In addition, a number of new methods have also been designed and have been described in references Ghoneim and Taha [11] – Li et al. [12] – Mansour [13].

2.3. Degradation model in reliability analysis

Failures have different causes which can be categorized either as external or internal. Internal failures happen due to internal system faults (e.g. ageing of materials). External failures are caused by environmental conditions (e.g. pollution, vibrations, and humidity). In general, failures can be divided into two groups: degradation failures and sudden failures. Degradation failures can be predicted by a single condition monitoring indicator or by several indicators. Furthermore, the degradation process effects and reduces the reliability of assets. Sudden failure cannot be predicted through condition monitoring techniques or the measurement of asset age. Figure 1 shows how assets reach the degradation level (failure threshold) and hence fail. Degradation has two types: natural degradation and forced degradation. Natural degradation occurs due to age, or is time-dependent. Forced degradation occurs due to external causes to the system (e.g. due to high demand of power, the load of power transformers increases significantly and reaches the failure threshold level and hence the transformer can no longer safely take the load) [14].

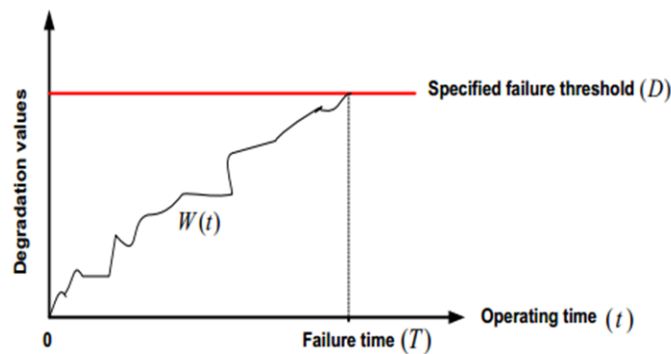


Figure 1. Degradation process. Source: [14]

2.4. Paper degradation of power transformer

Power systems have many different elements; the power transformer is the most expensive element and its faults can be very costly [11]. Power transformers are responsible for power transmission and distribution. The main function of transformers is to step-up or step-down voltage levels of substations. Negative economic impacts in electrical networks are often caused by unplanned outage of power transformers. It is important to determine any problems from an early stage, before disastrous failure occurs. Accurate identification of problems supports reliable power transformer operation. Alongside predictive and corrective maintenance, preventive maintenance may be considered more important, and it should be implemented to guarantee as far as possible the highest performance and reliability of power transformers. Condition monitoring in electrical networks provides useful data to maintenance staff to support activities such as: the planning of operation and maintenance activities; assessments of condition of equipment; predicting the remaining life of electrical assets, and discovering areas of further development and improvement. Power transformers have two windings: high voltage winding and low voltage winding, shown in figure 2. The material used in the insulation system of these copper windings is Kraft paper immersed in mineral oil. The aging transformer under the load can lead to oil and paper insulation degrades. The result of this situation will lead to catastrophic failure in electrical networks.

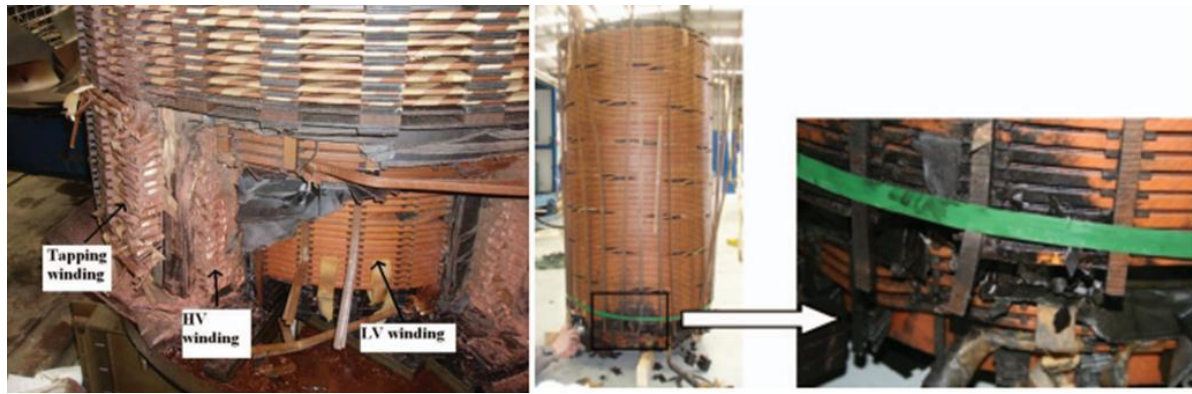


Figure 2. Failure of paper insulation of power transformer. Source: [15]

The power transformer tank contains paper, and its function is to insulate the two windings (high and low). This paper is made from a cellulosic material such as wood pulp. Over a period of time, high temperatures, water, and oil acids could potentially degrade the paper and hence the windings of the transformer could eventually fail. This paper has polymer chains and glucose rings which are linked together and repeated. Moreover, monomer is shaped from the repeating unit of glucose and has an average number of glucose per chain; this is known as the degree of polymerization (DP). Figure 3 shows how the paper is produced from long-chain molecules.

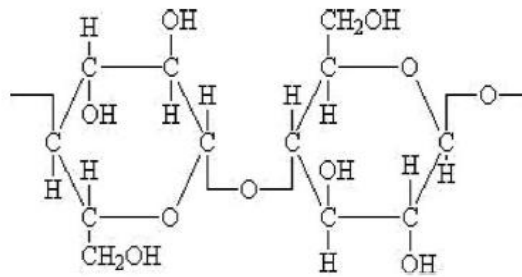


Figure 3. Cellulose Chain. Source: [15]

The state of paper insulation is very strong (electrical and mechanical strength) when the DP value of the power transformer is 1000 and above. On other hand, when the DP value is around 200 or less, it reflects the weak condition of paper insulation and the paper is unable to perform its required functions; therefore the power transformer could fail within a short period of time. Furthermore, there are many factors that can affect paper insulation age: high temperatures (pyrolysis), moisture around the paper (hydrolysis), and acid and oxygen (oxidation) within the transformer oil [15]. Over many decades, oil-paper insulation has been used inside transformer tanks due to its good reliability and performance. However, over a long period of operation, transformer paper could degrade as a result of many factors (its ageing, rising energy consumption, electrical and thermal stresses). These factors therefore could result in a reduction in the DP value as well as a reduction in electrical and mechanical strength of power transformers [16]. Paper insulation is therefore an important element in power transformers and its DP value gives an indication of when there are problems within internal parts. Condition monitoring of paper insulation, including its age (cellulose) helps to predict the remaining useful life of power transformers, and also identifies problems in their early stages, hence increasing transformer availability and reliability. In this way, asset managers can make informed decisions when authorising optimum asset replacements [17]. Figure 2 shows failure of paper insulation of power transformer.

2.5. Prognostic techniques for predicting remaining useful life

Prognostic is a predictive method that defines an asset's remaining useful life (RUL). It also estimates the time to failure and failure modes in the future. Failure prognosis includes predicting asset degradation based on condition monitoring techniques [18]. Figure 4 shows the fundamentals of prognostic techniques for degradation failure.

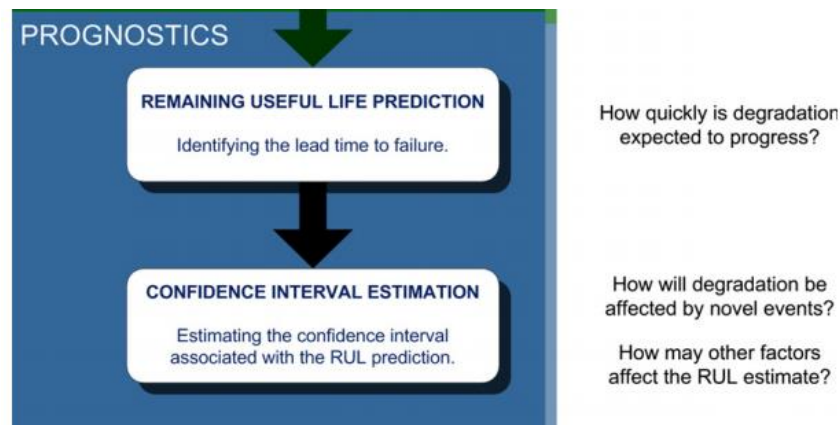


Figure 4. Prognostic process. Source: [18]

RUL has four main models that can be seen in figure 5; each model has a sub-model. Knowledge-based models assess the comparison between an observed condition and a database (failure event) of previous events in order to assume the life expectancy. It has two systems: expert and fuzzy systems. Life expectancy models define the life of single equipment components as well as determining expected risks from equipment deterioration. There are two models: stochastic models (Aggregate reliability functions and Conditional probability methods) and statistical models (Trend extrapolation, Auto-regressive moving average, and Proportional hazards modelling). Artificial neural network models calculate estimates for the remaining life of machines based on a mathematical representations from observation data. It has two models: direct RUL prediction, and estimation of parametrics for other models. Physical models calculate the estimation output for the remaining life of machines based on a mathematical representation of degradation processes [18].

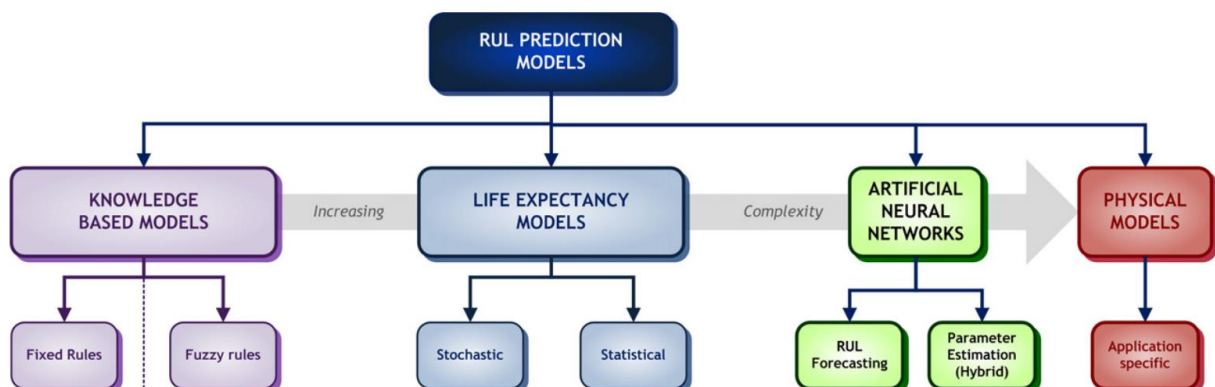


Figure 5. Main model categories for prediction of remaining useful life. Source: [18]

2.6. The P-F curve

Predictive technologies can be used in maintenance scheduling to build an essential PdM program. The P-F curve is significant for industrial assets; understanding the real meaning of this curve will lead to the implementation of appropriate maintenance strategies. PdM programs are therefore important during the P-F interval in order to detect component failures. It facilitates decisions to replace or repair parts of systems before faults occur, hence optimizing maintenance costs. Figure 6 illustrates how PdM can detect problems early during the P-F interval by using predictive technologies linked with predictive maintenance.

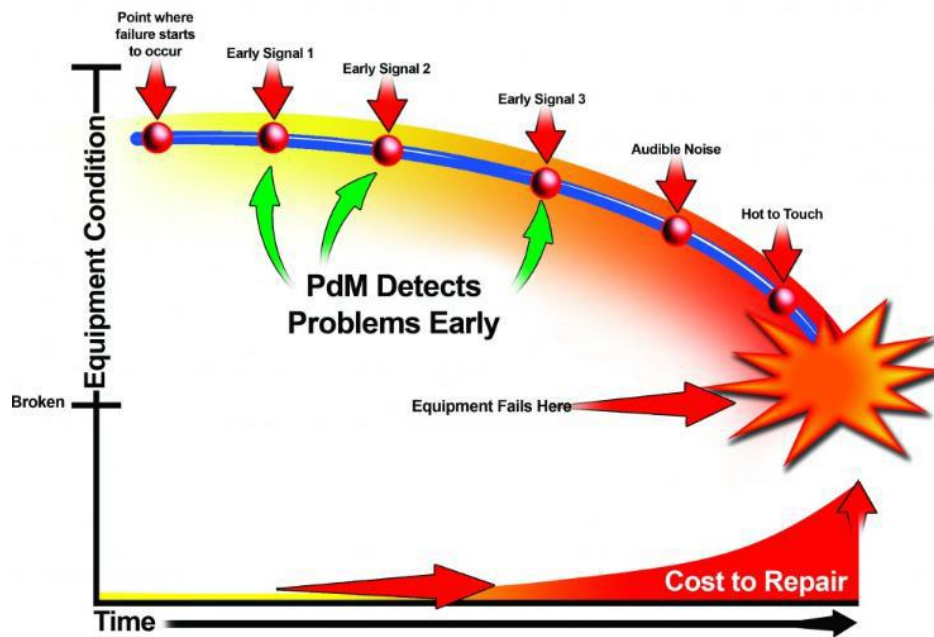


Figure 6. How PdM Detects Problems Early During P-F Interval. Source: [19]

3. RESEARCH METHODOLOGY

The needs and significance of TPM is well understood in many different industries. However, there has been almost no research applied in this area within Middle-East power plants. Thus, this research aims to examine the potential to implement TPM in the UAE power industry and determine if prognostic techniques can be critical success factor in planned maintenance pillar. A power organisation in Abu Dhabi power industry (AADC) was used as a case study to source data for the study to determine the successful TPM implementation [20]. Historical data from power transformers for five critical power distribution substations was collated. Power transformer paper degradation DP values were taken from condition monitoring data for a five year period, to perform required simulation analysis of prognostic techniques. The analysis required the application of ReliaSoft's Weibull++ 9 software in order to estimate the potential failure time of power transformers and hence select the correct maintenance policy based on a TPM strategy. Before the simulation analysis was applied it was essential to first evaluate the validity and reliability of the instrument to confirm that the findings would be valid and reliable.

4. RESULTS AND DISCUSSIONS

As illustrated in figure 7, ReliaSoft's Weibull++ 9 software is shown the prediction of warning limit (potential failure time in month) of power transformers paper insulation for five critical substations at AADC. This warning limit is equal to the "P" point in the P-F curve and its prediction is based on a

prognostic technique (Statistical Trend Extrapolation). Further, these paper insulations degrade over a long period of time due to poor maintenance practices and when they reach the functional failure (its DP value at less than 200). In this situation, the power transformers are replaced. Due to the power transformer renewal process and constant failure rate of paper insulation, the exponential distribution is adequate for this kind of failure mode (degradation of paper insulation). Therefore, exponential distribution is selected here to model the degradation data of paper insulation. After running the simulation, the Weibull++ 9 generates the potential failure predication time (P point in month) of the paper insulation for five critical power transformers at AADC and can be seen in table 1.

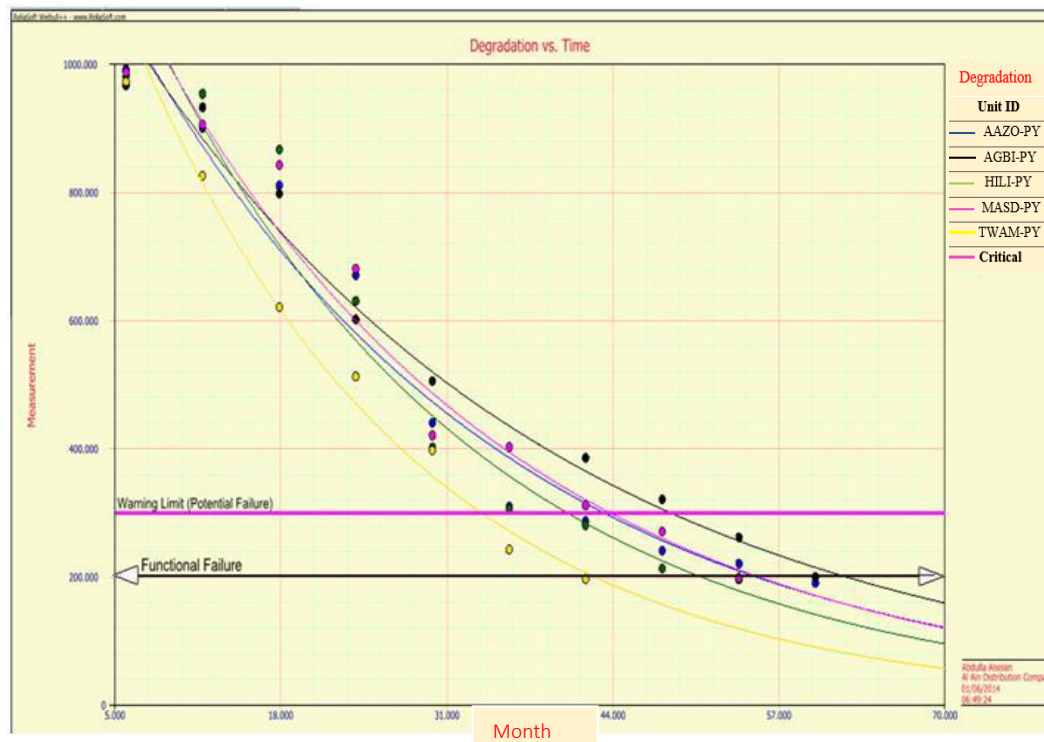


Figure 7. Estimating the P-F Interval of Power Transformer Paper Insulation at AADC.

Table 1: Plant failures by year

| Quick Results Report | | | |
|----------------------|---------------------------------------|-------------|---------|
| Report Type | Extrapolated Failure/Suspension Times | | |
| User Info | | | |
| User | Abdulla Alseiri | | |
| Company | Al Ain Distribution Company | | |
| Date | 1/6/2014 | | |
| Times | | | |
| | F/S | Time to F/S | Unit ID |
| | F | 43.28137416 | AAZO-PY |
| | F | 48.513284 | AGBI-PY |
| | F | 40.41985555 | HILI-PY |
| | F | 43.74875214 | MASD-PY |
| | F | 33.69817705 | TWAM-PY |

Successful electrical power companies have sufficient failure data that involves accurate time to failure of their power assets based on condition monitoring data. These data can estimate the potential failure time of their assets and hence select correct maintenance strategies. Warning limits (critical degradation) need to be selected carefully because if critical degradation is too high, it could cause failures to occur before they are predicted [18]. Thus, planned maintenance pillar of TPM can play a significant role by utilising predictive maintenance tools such as prognostic techniques and analysis through autonomous maintenance (operators). However, they should have the right skills, training and capability to use prognostic tools as

well interpret the historical data of power plants, and hence establish proper maintenance tasks to overcome any technical problems in early stages. Operator needs to understand the function, structure and inspection methods for each part of equipment [21] – [22]. This research has recognised how effective prognostic techniques can lead to successful implementation of TPM through planned maintenance pillar, as well highlighting that operators should collect, record and analyse data of power plants in order to provide benchmarks. Kodali et al. in [23] elaborated that the TPM framework advocates the use of predictive maintenance as it was promoted as one of the main pillar of Nakajima's framework. Maintenance costs involve direct and indirect expenses; direct costs related to materials and labour and indirect associated with six big losses [21] – [24]. According to the prognostic techniques utilised by operators, it will allow maintenance engineers and operators to use proposed maintenance tasks to create effective solutions and optimise maintenance costs [24] – [25]. Based on experience and identified potential failure during prognosis analysis, suitable tasks can be selected from maintenance policies and producers. Persons (operators or maintenance engineers) to perform tasks are identified according to complexity, and maintenance costs are calculated. This determines appropriate actions and resources (labour, tools, parts or instruments) required to implement the maintenance policies. The case study organisation has now an understanding of the crucial role of prognosis techniques and they have developed a basic maintenance task list which is placed next to the power plant.

5. CONCLUSION

The research has revealed that electric power organisations have recognised a strongly held perception that a reactive maintenance system works when the equipment is repaired at the same time of failure. However, the cost of fixing equipment in reactive mode can be three times greater than preventive maintenance costs. It seems that inherent maintenance practices, which reinforce reactive maintenance, also affect any changes towards implementation of world class maintenance practices. Furthermore, an organisation aspiring to employ world-class maintenance such as TPM will need to build a robust teamwork culture and utilise predictive maintenance concepts and tools such as prognostics techniques in order to predict the failures before occurring. The current practices within the Abu Dhabi power companies are static, not dynamic, and are underpinned by many obstacles and barriers. The results of this research have shown that effective prognostic techniques can lead electric power companies in Abu Dhabi to develop a TPM approach through planned maintenance, autonomous maintenance, training and education on the job, and focused improvement pillars. In addition, the results have proven that collection and analysis of historical data through operators is possible by using effective prognostic tools. However, implementing this new maintenance strategy (TPM) in the Abu Dhabi power industry may be challenging. Whilst TPM strategies will allow operators to conduct maintenance tasks and prognosis analysis, maintenance engineer responsibilities will need to be changed to enhance and support equipment operators. These changes will significantly affect the responsibilities and roles of operation staff. This research confirms that prognostic techniques can have a positive impact towards successful TPM implementation in the power industry.

REFERENCES

- [1] Reiman, T., Oedewald, P. (2004) Measuring maintenance culture and maintenance core task with CULTURE-questionnaire—a case study in the power industry. *Safety Science*. vol. 42, no. 9, pp. 859-889.
- [2] Nilsson, J., Bertling, L. (2007) Maintenance Management of Wind Power Systems Using Condition Monitoring Systems-Life Cycle Cost Analysis for Two Case Studies. *IEEE Transactions on Energy Conversion*. vol. 22, no. 1, pp. 223-229.
- [3] Bana E Costa, Carlos A., Carnero, M.C., Oliveira, M.D. (2012) A multi-criteria model for auditing a predictive maintenance programme. *European Journal of Operational Research*. vol. 217, no. 2, pp. 381-393.
- [4] Nakajima, S. (1988) *Introduction to TPM*. Productivity Press, Cambridge.
- [5] Labib, A.W. (1999) A frame work for benchmarking appropriate productive maintenance. *Management Decision*. vol 37, no.10, pp 792-799.
- [6] Lee Cooke, F. (2000) Implementing TPM in plant maintenance: some organisational barriers. *International Journal of Quality and Reliability Management*. vol. 17, no. 9, pp. 1003-1016.
- [7] Ahuja I.P.S., Khamba, J.S. (2008) Total productive maintenance: Literature review and directions. *International Journal of Quality and Reliability Management*. vol.25, no.7, pp. 709-756.

- [8] Waeyenbergh, G., Pintelon, L. (2002) A framework for maintenance concept development. *International Journal of Production Economics*. vol. 77, no. 3, pp. 299-313.
- [9] Schmidt, B., Wang, L. (2016) Cloud-enhanced predictive maintenance. *The International Journal of Advanced Manufacturing Technology*. ISSN:0268-3768.
- [10] Malik, H., Mishra, S. (2016) Application of Gene Expression Programming (GEP) in Power Transformers Fault Diagnosis Using DGA. *IEEE Transactions on Industry Applications*. vol. 52, no. 6, pp. 4556-4565.
- [11] Ghoneim, S.S.M., Taha, I.B.M. (2016) A new approach of DGA interpretation technique for transformer fault diagnosis. *International Journal of Electrical Power and Energy Systems*. vol. 81, pp. 265-274.
- [12] Li, X., Wu, H., Wu, D. (2011) DGA Interpretation Scheme Derived From Case Study. *IEEE Transactions on Power Delivery*. vol. 26, no. 2, pp. 1292-1293.
- [13] Mansour, D.A. (2015) Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases. *IEEE Transactions on Dielectrics and Electrical Insulation*. vol. 22, no. 5, pp. 2507-2512.
- [14] Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., Sun, Y. (2009) A review on degradation models in reliability analysis. *Engineering Asset Lifecycle Management - Proceedings of the 4th World Congress on Engineering Asset Management*. pp. 369.
- [15] Lelekakis, N., Guo, W., Martin, D., Wijaya, J., Susa, D. (2012) A field study of aging in paper-oil insulation systems. *IEEE Electrical Insulation Magazine*. vol. 28, no. 1, pp. 12-19.
- [16] Wang, S., Zhang, G., Mu, H., Wang, D., Lei, M., Suwarno, S., Tanaka, Y., Takada, T. (2012) Effects of paper-aged state on space charge characteristics in oil-impregnated paper insulation. *IEEE Transactions on Dielectrics and Electrical Insulation*. vol. 19, no. 6, pp. 1871-1878.
- [17] Baird, P.J., Herman, H., Stevens, G.C., Jarman, P.N. (2006) Non-destructive measurement of the degradation of transformer insulating paper. *IEEE Transactions on Dielectrics and Electrical Insulation*. vol. 13, no. 2, pp. 309-318.
- [18] Sikorska, J.Z., Hodkiewicz, M., Ma, L. (2011) Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*. vol. 25, no. 5, pp. 1803-1836.
- [19] Gehloff, M. (2013) P-F Curve 101- The Basics Maintenance Phoenix. Available: <http://maintenancephoenix.com/2013/10/02/pf-curve-101-keeping-it-simple/>.
- [20] Yin, R.K. (2014) Case study research: design and methods. SAGE, Los Angeles.
- [21] Baglee, D., Knowles, M. (2010) Maintenance strategy development within SMEs: The development of an integrated approach. *Control and Cybernetics*. vol. 39, no. 1, pp. 275-303.
- [22] Bamber, C.J., Sharp, J.M., Hides, M.T. (1999) Factors affecting successful implementation of total productive maintenance: a UK manufacturing case study. *Journal of Quality in Maintenance Engineering*. vol 5, no.3, pp. 162-81.
- [23] Kodali, R., Prasad Mishra, R., Anand, G. (2009) Justification of world-class maintenance systems using analytic hierarchy constant sum method. *Journal of Quality in Maintenance Engineering*. vol. 15, no. 1, pp. 47-77.
- [24] Moubray, J. (1997) Reliability-centred maintenance. Butterworth-Heinemann, Oxford.
- [25] Singh, J., Sood, Y.R., Jarial, R.K. (2008) Condition Monitoring of Power Transformers - Bibliography Survey. *IEEE Electrical Insulation Magazine*. vol. 24, no. 3, pp. 11-25.

Fault Diagnosis of a Polymer Electrolyte Membrane Fuel Cell using Bayesian Network

L. Mao¹, L. Jackson¹, B. Davies¹

¹ Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3RX, UK

ABSTRACT

In this paper, a Bayesian network of a polymer electrolyte membrane (PEM) fuel cell is developed for fault diagnosis, which consists of several fuel cell failure modes relating to the fuel cell components via various failure mechanisms. In order to better utilize the Bayesian network in practical PEM fuel cells, an extra sensor layer is added to the Bayesian network, as the variation in sensors can be obtained directly from fuel cell tests. The relationship between added sensors and fuel cell failure modes can be determined based on previous studies, while their conditional probabilities can be calculated using fuel cell degradation rates due to different failure modes. To validate the effectiveness of developed the fuel cell Bayesian network, a case study of fuel cell flooding is investigated, the fault diagnosis is carried out by monitoring the variation in the sensors. Results demonstrate that the flooding can be successfully identified by monitoring the sensor measurements from the PEM fuel cell system.

Keywords: PEM fuel cell, system fault diagnosis, Bayesian network .

Corresponding author: Dr L. Mao (l.mao@lboro.ac.uk)

1. INTRODUCTION

In the last few decades, with rapid development of hydrogen and fuel cell technologies, especially polymer electrolyte membrane (PEM) fuel cells with characteristics like zero-emissions and high efficiency, they have been equipped in several applications including stationary power station and automotive. However, the reliability and durability of fuel cells during their lifetime are still two major barriers for their further commercialization.

To address these issues, fault diagnostic techniques have been applied to fuel cells to detect and isolate the fuel cell faults. Based on the results, mitigation strategies can be taken to extend the fuel cell lifetime. From the literatures, the fuel cell fault diagnostic techniques can be divided into two groups, model-based [1-3] and data-driven methods [4-6]. Compared to model-based techniques, more studies have utilized data-driven approaches for fuel cell fault diagnosis. This is mainly due to the difficulty of developing an accurate fuel cell model with inclusion of complete failure mode effects.

However, one major drawback of data-driven fault diagnostic techniques is that the diagnostic performance relies heavily on the quantity and quality of fuel cell data, i.e. the inclusion of measurement noise and data not sensitive to fuel cell performance variations may affect the diagnostic results.

To improve the diagnostic performance and alleviate the test data dependency, the relationships between fuel cell components and expert knowledge about failure mechanisms should be included in the diagnosis. Bayesian network is such a probabilistic graphical model representing a set of variables in a system and their conditional dependencies, and can be used to deal with uncertain and incomplete information from the PEM fuel cell system.

Some researchers have performed fuel cell fault diagnosis using Bayesian networks [7-8], however, among these studies, limited fuel cell failure modes are considered in the construction of the Bayesian network,

thus the developed Bayesian network can only be used to identify specific fuel cell failure modes. Therefore, it is highly desirable to construct a fuel cell Bayesian network including complete failure modes, which can be used for practical PEM fuel cell fault diagnosis.

This paper presents the development of fuel cell Bayesian network with complete failure modes, and applies the Bayesian network for PEM fuel cell fault diagnosis. In section 2, the PEM fuel cell system is briefly described. Section 3 describes the construction of the fuel cell Bayesian network and determination of the conditional probabilities between the components. In section 4, the flooding scenario is created from a PEM fuel cell system to evaluate the effectiveness of developed Bayesian network. From the results, conclusions are given in section 5.

2. DESCRIPTION OF PEM FUEL CELL SYSTEM

A typical PEM fuel cell shown in figure 1 includes two bipolar plates, and they are separated by a membrane electrode assembly (MEA), which consists of a polymer electrolyte membrane, electrodes, catalyst and gas diffusion layers.

During operation, hydrogen and air/oxygen are injected into the anode and cathode, respectively. At the anode side, hydrogen is oxidised to release protons and ions, where protons can pass through the membrane directly, and ions can only reach the cathode side via the external circuit. At the cathode side, protons, ions, and oxygen can react to produce heat and water, which can be removed from the cathode outlet, these reactions can be written as:

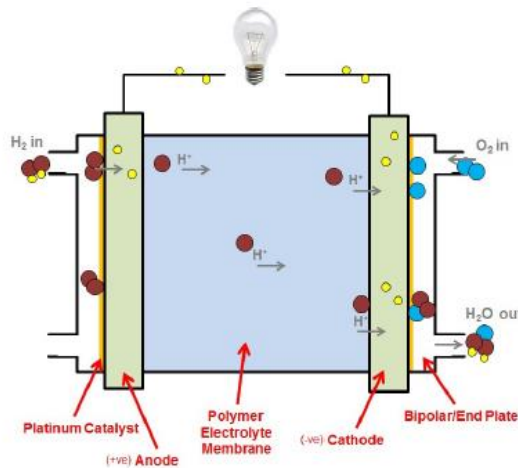
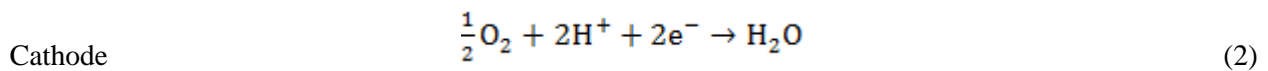
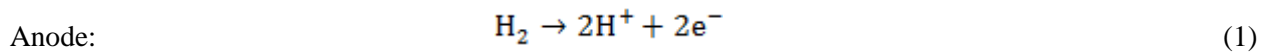


Figure 1. Typical PEM fuel cell stack [9]

In a practical fuel cell system, a certain number of PEM fuel cells are usually connected in series to form a fuel cell stack to provide the required output power. Several ancillaries, such as gas supply systems, cooling system and humidification system, should be used during the fuel cell system operation to provide gas reactants and control operating conditions.

3. CONSTRUCTION OF PEM FUEL CELL BAYESIAN NETWORK

Generally, a Bayesian network can be written in the following format:

$$B=(G,CP) \quad (3)$$

where B is the system Bayesian network, G is the directed acyclic graph (DAG) representing the system graphical structure, CP is the conditional probabilities capturing the probabilistic dependence among system variables.

From Eq.(3) it can be seen that the DAG and conditional probabilities are the two components when developing a system Bayesian network, thus in the following parts, the construction of the DAG and the determination of conditional probabilities for the PEM fuel cell will be presented.

3.1. Construction of PEM fuel cell Bayesian structure

From previous studies, there exist two groups of techniques for constructing DAG, including methods based on human knowledge about the process, and techniques based on probabilistic algorithms using databases of records, in which two typical methods are widely used, constraint-based and search-and-score [7]. In this study, since the fuel cell Bayesian network is constructed including complete fuel cell failure modes, it is difficult to obtain the records of databases for all the fuel cell failure modes, thus human knowledge obtained from previous studies is used herein for the construction of DAG.

Based on studies relating to PEM fuel cell fault tree and Bayesian network [7-11], the DAG of fuel cell Bayesian network is developed and depicted in figure 2, which consists of 4 layers and 12 failure modes listed in Table 1.

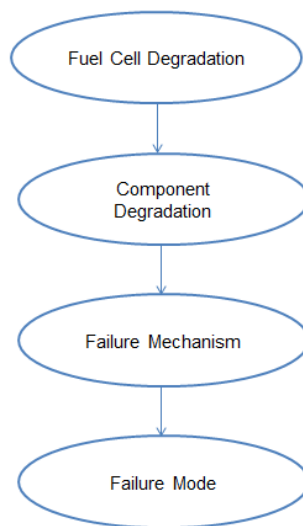


Figure 2. Structure of developed PEM fuel cell Bayesian network.

Table 1: PEM fuel cell failure modes in developed Bayesian network

| PEM fuel cell failure mode | | |
|--------------------------------------------------------|---------------------------|---------------------------------------------------------|
| Flooding | Excess heat | Pt loss, migration, agglomeration |
| Incorrect bipolar plate (BIP) torque | Creep | Exothermal combustion due to previously formed pinholes |
| Fatigue from relative humidity and temperature cycling | Dissolution of metal ions | Corrosion leading to release of multivalent cations |
| Contamination from humidifier/air pipe/gas impurity | OH and OOH radical attack | Previous formed pinholes |

However, a drawback of this structure is that the inputs to the developed PEM fuel cell Bayesian network are the states of fuel cell failure modes, which are also the target of fuel cell fault diagnosis, thus more information should be added to the Bayesian network in order to perform fault diagnosis.

To address the above issue, 6 sensors are added to link the failure modes, which are listed in table 2. With the link between sensors and failure modes shown in figure 3 [12-16], the fuel cell fault diagnosis can be performed straightforwardly, as sensor readings can be observed directly from the tests with cell under operation. With the performance degradation exceeding the predefined threshold, the analysis will be triggered to infer the probabilities of different fuel cell failure modes. It should be mentioned as these sensors are placed at the fuel cell outlet to capture the fuel cell outputs, the constructed Bayesian network can only be used in fuel cell steady state condition, i.e. the fuel cell inputs are constant during the system operation. Moreover, for simplification purposes, only 5 failure modes listed in table 3 are selected in the following analysis, the selection is made based on the fact that with only 6 sensors and 2 discrete values for each sensor, several failure modes will give the same effect, thus without addition of further sensor information, fewer failure modes are used which can cause different sensor variation and can be discriminated with 6 sensors.

Table 2: Sensors added to the developed Bayesian network

| Sensor | |
|-------------------------|-----------------------------|
| Fuel cell stack voltage | Fuel cell stack temperature |
| Anode outlet pressure | Cathode outlet pressure |
| Anode humidity | Cathode humidity |

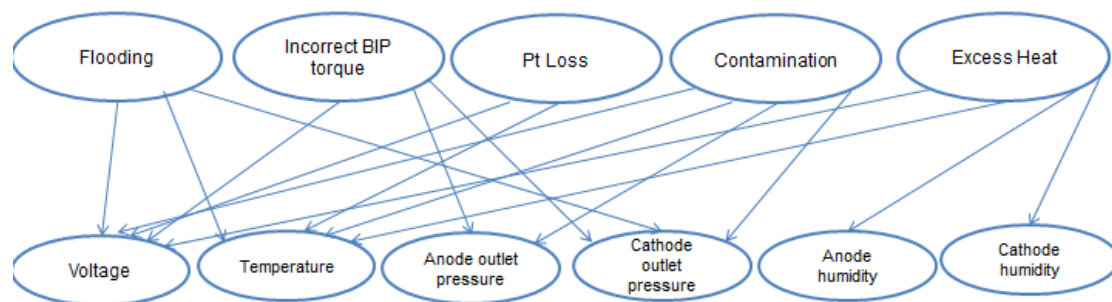


Figure 3. Relationship between fuel cell failure mode and sensor reading

Table 3: PEM fuel cell failure modes used in the current study

| PEM fuel cell failure mode | | |
|--------------------------------------|-----------------------------------------------------|-----------------------------------|
| Flooding | Excess heat | Pt loss, migration, agglomeration |
| Incorrect bipolar plate (BIP) torque | Contamination from humidifier/air pipe/gas impurity | |

3.2. Determination of conditional probability

In the current analysis, the discrete state is assigned to the variables in the PEM fuel cell Bayesian network. For each node in the Bayesian network, two discrete states are used (True and False).

Firstly, the probabilities of fuel cell failure modes are calculated based on previous studies [11-12], the reciprocal of the degradation rate (caused by the corresponding fuel cell failure mode) follows the Weibull distribution shown in Eq.(4), and by calculating scale and shape parameters of the Weibull function, the probability of the fuel cell failure mode within a certain period can be determined. Table 4 lists the scale and shape parameters for different failure modes [12], it should be mentioned that in this study, the fuel cell failure is defined when there is a 5% voltage drop.

$$F(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} \quad (4)$$

where β and η are the shape and scale parameters, respectively.

Table 4: Scale and shape parameters in Weibull distribution of different failure modes

| Failure mode | Scale parameter (η) | Shape parameter (β) |
|-----------------------------------------------------|----------------------------|-----------------------------|
| Flooding | 0.13 | 1 |
| Excess heat | 0.2 | 1 |
| Pt loss, migration, agglomeration | 1.77 | 0.8 |
| Incorrect bipolar plate (BIP) torque | 4.41 | 0.8 |
| Contamination from humidifier/air pipe/gas impurity | 12.91 | 2 |

With the relationship between failure modes and sensors depicted in figure 3, the conditional probabilities for the sensors can be determined. Table 5 shows an example of the conditional probability for the anode outlet pressure sensor. It should be noted that ‘true’ failure mode indicate the existence of the failure mode, while the ‘true’ sensor means the variation in the sensor exceeds the threshold value, which is described in the following section.

Table 5: Conditional probability of anode outlet pressure sensor

| Incorrect BIP torque | Contamination | Anode outlet pressure | |
|----------------------|---------------|-----------------------|------|
| | | False | True |
| False | False | 1 | 0 |
| False | True | 0 | 1 |
| True | False | 0 | 1 |
| True | True | 0 | 1 |

4. APPLICATION OF DEVELOPED BAYESIAN NETWORK IN PEM FUEL CELL FAULT DIAGNOSIS

As a typical PEM fuel cell failure mode, flooding can cause the fastest fuel cell performance degradation rate [9], while with the detection of fuel cell flooding, mitigation strategies, like increased purging rate, can be implemented to recover the fuel cell performance. Therefore, in this study, the PEM fuel cell flooding is used to evaluate the effectiveness of the developed Bayesian network.

4.1. PEM fuel cell experimental test

The single PEM fuel cell with 100cm² is tested in an 800W test bench shown in figure 4, and table 6 lists some technical information about the PEM fuel cell. In addition, several supporting systems, such as air/hydrogen supply systems, temperature control circuit, humidification system, are used to operate the PEM fuel cell system.



Figure 4. PEM fuel cell test bench

Table 6: PEM fuel cell parameters

| Parameter | Value |
|-------------------------|------------------------|
| Membrane thickness | 25μm |
| Active area | 100cm×100cm |
| Platinum loading | 0.2 mg/cm ² |
| Gas diffusion thickness | 415 μm |
| Reactant stoichiometry | 1/3 @ nominal (an/ca) |

In the test, the PEM fuel cell is firstly operated at nominal conditions for a certain duration to quantify sensor variation without performance degradation, which can be used to determine the threshold for fuel cell abnormal performance, then the fuel cell stack temperature is decreased to create water condensation. The performance degradation can be attributed to PEM fuel cell flooding due to the divergence between stack temperature and gas dew point.

Figure 5 depicts the sensor measurements at the PEM fuel cell nominal condition and the thresholds for the PEM fuel cell nominal condition, this is determined using Eq.(5) based on the fact that the all the sensor variations at nominal condition will not exceed the threshold value. It can be seen that the variation of sensor measurement at nominal conditions is within range defined with the threshold values, and when the fuel cell stack voltage exceeds the threshold, the fault diagnosis will be triggered.

$$\text{thres} = \text{mean}(S) \pm 3 \times \text{std}(S) \quad (5)$$

Where ‘thres’ is the threshold value, ‘S’ is the sensor readings, ‘mean’ and ‘std’ are the mean and standard deviation operations, respectively.

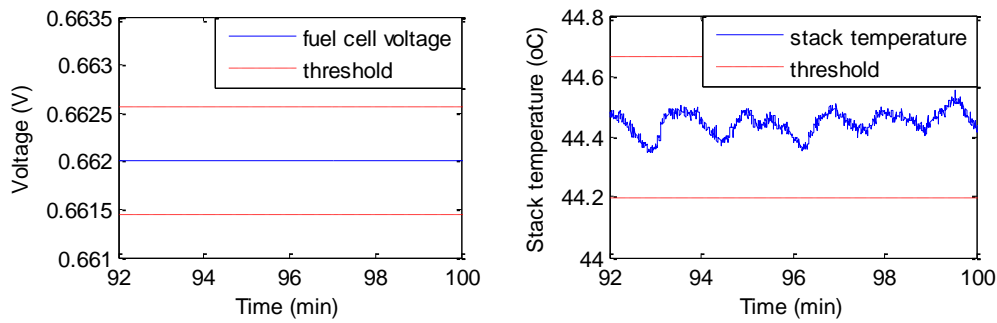


Figure 5. Fuel cell voltage, and temperature at nominal conditions

With the fuel cell flooding, the sensor measurements will be changed, which are depicted in figure 6. From the figure, clearly performance degradation can be observed due to flooding. Moreover, by further studying the sensor variations with defined threshold values from Eq.(4), it can be seen that there is a small time delay between stack temperature drop and the fuel cell voltage drop, indicating a certain time is required for the condensed water to cause fuel cell voltage drop.

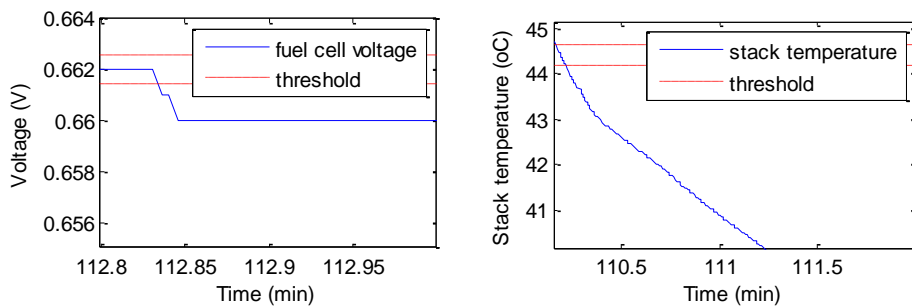


Figure 6. Fuel cell voltage and temperature variation at flooding scenario

With the stack voltage exceeding the predefined threshold, the fault diagnosis is carried out to infer the probabilities of different fuel cell failure modes. It should be mentioned that in the current test, variations in 3 sensors are observed (fuel cell voltage, stack temperature, and cathode outlet pressure), and the remaining 3 sensors are within the threshold range. Table 7 lists the inferred probable fault causes for the drop of these sensor readings.

Table 7: Probable fault causes for the observed sensor readings

| Fault | Probability |
|-----------------------------------|-------------|
| Flooding | 100% |
| Pt loss, migration, agglomeration | 77% |

It can be seen from table 7 that flooding is the most possible fault cause for the fuel cell performance degradation, while Pt loss, migration and agglomeration is the less possible fault cause. The reason is that Pt loss, migration and agglomeration will only cause variations in fuel cell voltage and temperature, which is also included in the consequences due to flooding. It is expected that with further inclusion of more sensors in the Bayesian network, the performance of the Bayesian network in PEM fuel cell fault diagnosis will be improved.

5. CONCLUSION

This paper presents the development of the PEM fuel cell Bayesian network, and the effectiveness of the Bayesian network in PEM fuel cell fault diagnosis. Compared to the previous developed PEM fuel cell Bayesian network, more fuel cell failure modes are added in the Bayesian network, and several sensors are added to facilitate the fault diagnosis, as the sensor readings can be accessed directly from the fuel cell test, the fault diagnosis can be triggered by studying the variation of sensor readings. The developed Bayesian network is then applied to PEM fuel cell system to identify the flooding scenario, where results demonstrate that the inference from the Bayesian network matches the original fault cause. This paves the way of using the developed Bayesian network for on-line fault diagnosis of PEM fuel cell system. Further study will be carried out to validate the whole PEM fuel cell Bayesian network and add more sensors for better fault diagnostic performance.

ACKNOWLEDGEMENT

This work is supported by grant EP/K02101X/1 for Loughborough University, Department of Aeronautical and Automotive Engineering from the UK Engineering and Physical Sciences Research Council (EPSRC).

REFERENCE

- [1] R. Petrone, Z. Zheng, D. Hissel, M.C. Pera, C. Pianese, M. Sorrentino, M. Becherif, N.Y. Steiner. (2013) A review on model-based diagnosis methodologies for PEMFCs, *International Journal of Hydrogen Energy* 38 7077-7091.
- [2] N. Fouquet, C. Doulet, C. Nouillant, G.D. Tanguy, B.O. Bouamama. (2006) Model based PEM fuel cell state-of-health monitoring via ac impedance measurements, *Journal of Power Sources* 159 905-913.
- [3] M.M. Kamal, D. Yu. (2011) Model-based fault detection for proton exchange membrane fuel cell systems, *International Journal of Engineering Science and Technology* 3 1-15.
- [4] Z. Zheng, R. Petrone, M.C. Pera, D. Hissel, M. Becherif, C. Pianese, N.Y. Steiner, M. Sorrentino. (2013) A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems, *International Journal of Hydrogen Energy* 38 8914-8926.
- [5] E. Pahon, N.Y. Steiner, S. Jemei, D. Hissel, P. Mocoteguy. (2016) A signal-based method for fast PEMFC diagnosis, *Applied Energy* 165 748-758.
- [6] L. Mao, L. Jackson, S.J. Dunnett. (2016) Fault diagnosis of practical polymer electrolyte membrane (PEM) fuel cell system with data-driven approaches, *Fuel Cells* DOI: 10.1002/fuce.201600139.
- [7] L.A.M. Riascos, M.G. Simoes, P.E. Miyagi. (2007) A Bayesian network fault diagnostic system for proton exchange membrane fuel cells, *Journal of Power Sources* 165 267-278.
- [8] L.A.M. Riascos, M.G. Simoes, P.E. Miyagi. (2008) On-line fault diagnostic system for proton exchange membrane fuel cells, *Journal of Power Sources* 175 419-429.
- [9] M. Whiteley, S. Dunnett, L. Jackson. (2016) Failure mode and effect analysis, and fault tree analysis of polymer electrolyte membrane fuel cells, *International Journal of Hydrogen Energy* 41 1187-1202.
- [10] A. Benmouna, M. Becherif, D. Depernet, F. Gustin, H.S. Ramadan, S. Fukuhara. (2017) Fault diagnosis methods for proton exchange membrane fuel cell system, *International Journal of Hydrogen Energy* 42 1534-1543.
- [11] L. Placca, R. Kouta. (2011) Fault tree analysis for PEM fuel cell degradation process modelling, *International Journal of Hydrogen Energy*, 36 12393-12405.
- [12] M. Whiteley. (2016) Advanced reliability analysis of polymer electrolyte membrane fuel cells in automotive applications, PhD thesis, Loughborough University, UK.
- [13] R. Borup, J. Meyers, B. Pivovar, Y.S. Kim, R. Mukundan, N. Garland, D. Myers, M. Wilson, F. Garzon, D. Wood, P. Zelenay, K. More, K. Stroh, T. Zawodzinski, J. Boncella, J.E. McGrath, M. Inaba, K. Miyatake, M. Hori, K. Ota, Z. Ogumi, S. Miyata, A. Nishikata, Z. Siroma, Y. Uchimoto, K. Yasuda, K. Kimijima, N. Iwashita. (2007) Scientific aspects of polymer electrolyte fuel cell durability and degradation, *Chem. Rev.* 107 3904-3951.
- [14] P. Rama, R. Chen, J. Andrews. (2008) A review of performance degradation and failure modes for hydrogen-fuelled polymer electrolyte fuel cells, *Proceedings of the institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 222 421-441.
- [15] J. Wu, X.Z. Yuan, J.J. Martin, H. Wang, J. Zhang, J. Shen, S. Wu, W. Merida. (2008) A review of PEM fuel cell durability: degradation mechanisms and mitigation strategies, *Journal of Power Sources*, 184 104-119.
- [16] W. Schmittinger, A. Vahidi. (2008) A review of the main parameters influencing long-term performance and durability of PEM fuel cells, *Journal of Power Sources*, 180 1-14.

Weighted Narrowband Envelope Spectrum and its Application to Bearing Fault Diagnosis

Jie Duan¹, Tielin Shi¹, Jianping Xuan¹, Hongdi Zhou¹

¹ School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, Peoples R China

ABSTRACT

The Protrugram [1] is based on the kurtosis of the envelope spectrum amplitudes of the narrowband demodulated signal, and has been proved practical and powerful in bearing fault diagnosis. However, kurtosis is very sensitive to large random impulses that are frequently encountered in industrial applications. So the narrowband with maximum kurtosis may not be associated with the bearing fault. Meanwhile, the key issue of bearing fault diagnosis is to detect the bearing's characteristic frequencies rather than eliminate all noises. So a novel method is presented by combining the envelope spectra at different frequency bands for the bearing fault diagnosis in this paper. The Jarque–Bera statistic, a combination of kurtosis and skewness, is used to detect the existence of impulses in the envelope spectrum. These narrowband envelope spectra with larger Jarque–Bera statistics are selected, and weighted mean of the sequence of envelope spectra is calculated as the weighted narrowband envelope spectrum. Then the weighted spectrum can be used to determine the types of bearing faults by identifying its characteristic frequencies. The effectiveness of the proposed method in bearing fault detection is validated using some real signals.

Keywords: bearing fault diagnosis, narrowband amplitude demodulation, Jarque–Bera statistic.

Corresponding author: Tielin Shi (email: tlshi@mail.hust.edu.cn)

1. INTRODUCTION

Bearings are among the most common machine components to be found in rotating machinery. Bearing failure causes about 45-55% of rotating equipment failures in the industry [2]. Therefore, detection and diagnosis of incipient bearings faults are becoming important to prevent malfunction and failure. When a rolling element moves over the damage position on the bearing, an impulse of short duration will be generated, which would excite resonance of the bearing and adjacent components in the machine [3]. These impulses occur periodically at characteristic defect frequency, determined by the defect location of bearings. Hence, it is possible to identify the types of bearing faults by extracting the impulsive features in the vibration signals.

Envelope analysis (EA) is a demodulation method and has been used for extracting periodic shocks in vibration signals. It has been widely used in bearing fault detection in recent years. The signals are often contaminated by the noise present in the measurement device and the interferences generated by the vibrations of other machinery components. Thus, band-pass filter is often used to wipe off these out-band interferences before demodulation is performed. The key point of this technique is determining the resonance frequency and bandwidth. Spectral kurtosis (SK), originally presented by Dwyer [4], was used to discover the presence of non-Gaussian components and indicate in which frequency bands these components occur. Subsequently, Antoni [5] [6] made a thorough and interesting study on SK and proposed Kurtogram [7]. The Kurtogram is based on the short time Fourier transform (STFT) or FIR filters, and has been widely used in bearing fault diagnosis [8-10]. However, temporal signal-based kurtosis can be considerably affected by non-Gaussian noise containing high peaks. Hence, a new method named Protrugram which takes advantage of the kurtosis of envelope spectra amplitudes as a function of the center frequency was proposed in [1]. The kurtosis of envelope spectrum amplitudes is an efficient index for band

selection [11]. It is logical to measure kurtosis in the frequency domain. Nevertheless, kurtosis is easy to be affected by large random impulses that are frequently encountered in industrial applications. The narrowband with maximum kurtosis may not be associated with the bearing fault. In this work, a weighted narrowband envelope spectrum obtained by combining the envelope spectra at multiple frequency bands is proposed for the bearing fault diagnosis.

The paper is organized as follows: Section 2 presents the realize process of weighted narrowband envelope spectrum in bearing faults diagnosis. In Section 3, we used real data obtained in the laboratory to validate the proposed method. Conclusions are summarized in Section 4.

2. WEIGHTED NARROWBAND ENVELOPE SPECTRUM

Recently, Barszcz and Jabłoński [1] proposed a novel method called Protrugram to select the optimal frequency band for the amplitude demodulation. The technique takes advantage of kurtosis values of the narrowband envelope spectral amplitudes, and has been proved efficient in bearing faults diagnosis even in the low signal-to-noise ratio situation [1] [11] [12]. It is based on the assumption that the harmonics of the bearing defect frequency are determined impulses in the envelope spectrum while the noises are randomly distributed over the whole frequency band. When the measuring environment is relatively clear, the hypothesis is still tenable. However, there are many interference factors in the industrial field, which make the vibration signals more complicated. The narrow band with maximum kurtosis may not be associated with the bearing fault in the situation. Meanwhile, the key issue of bearing fault diagnosis is to detect the bearing's characteristic frequencies rather than eliminate all noises and interferences. So the weighted narrowband envelope spectrum (WNES) is proposed in this work. It makes use of weighted averaging of a sequence of envelope spectra at different bands. These bands are selected based on Jarque–Bera (*JB*) statistic [13] rather than kurtosis as it has a better performance for the impulses detection in the spectrum. The scheme diagram of the proposed method is shown in figure 1 and the details are described below.

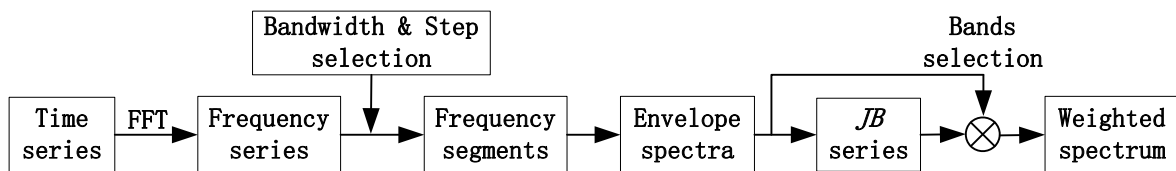


Figure 1. Scheme diagram of the proposed method.

Step 1. The vibration signal measured by an accelerometer is transformed to frequency domain at first. The signal can be expressed as a time series $x(k)$, where $k=1,2,\dots,N$, and N is the length of the series. Then the signal can be transformed to frequency domain by Fourier transform:

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{-j(2\pi/N)kn} \quad (1)$$

where $X(*)$ is the frequency series of the signal.

Step 2. The frequency series are split into segments. This process can be expressed as a frequency window slips over the frequency domain and truncates the frequency series [14]. Window width and step length are two important parameters. The frequency window width is the same as bandwidth in Protrugram. And the window can be defined as

$$W(k) = \begin{cases} 1 & 0 \leq k < N_w \\ 0 & \text{else} \end{cases} \quad (2)$$

N_w is the window length, which is chosen slightly more than three times the sought characteristic frequency as the same in the Protrugram. Then the frequency segments can be calculated as

$$X_i(n) = X(n) \cdot W(n - (i-1)a) \quad (3)$$

where X_i is the i th segment, a is shift step length. So the segments series can be obtained as follows:

$$S = \{X_1, X_2, \dots, X_m\} \quad (m = \lceil (N - N_w) / a \rceil) \quad (4)$$

The symbol $\lceil \rceil$ in the formula gets the maximum integer which is not more than the operated data. The step length a and window length N_w should be carefully picked because of the significant influence on S .

Step 3. For each segment, the JB statistic of the envelope spectrum amplitudes is calculated. The frequency segment is transformed to time domain by IFFT at first. Then the envelope can be extracted by Hilbert transform and other simple calculations. The narrowband envelope spectrum is obtained by Fourier transforming of the envelope. The spectrum sequence is defined as

$$SE = \{E_1, E_2, \dots, E_m\} \quad (5)$$

where E_i is the envelope spectrum of the i th segment. JB statistic is a combination of empirical skewness and kurtosis, and has been proved to be more effective than kurtosis in informative frequency band selection. For each segment, JB statistic is calculated as

$$JB_i = \frac{N}{6} \left[S_i^2 + \frac{(K_i - 3)^2}{4} \right] \quad (6)$$

where S_i and K_i are the skewness and kurtosis of the series E_i , which are defined as

$$S_i = \frac{\left((1/N) \sum_{j=0}^{N-1} [E_i(j) - ((1/N) \sum_{j=0}^{N-1} E_i(j))] \right)^3}{\left\{ \left((1/N) \sum_{j=0}^{N-1} [E_i(j) - ((1/N) \sum_{j=0}^{N-1} E_i(j))] \right)^2 \right\}^{3/2}} \quad (7)$$

and

$$K_i = \frac{\left((1/N) \sum_{j=0}^{N-1} [E_i(j) - ((1/N) \sum_{j=0}^{N-1} E_i(j))] \right)^4}{\left\{ \left((1/N) \sum_{j=0}^{N-1} [E_i(j) - ((1/N) \sum_{j=0}^{N-1} E_i(j))] \right)^2 \right\}^2} \quad (8)$$

After calculate JB statistic of each segment, JB statistic series can be obtained as $\{JB_1, JB_2, \dots, JB_m\}$.

Step 4. Weighted narrowband envelope spectrum is acquired by weighted averaging of envelope spectra at selected frequency bands. JB statistic can indicate the fault related harmonics presented in the envelope

spectrum. So the mean of JB statistic series JB_I is calculated as the threshold for bands selection. These narrowband spectra with JB values bigger than JB_I are selected for averaging. Suppose p bands are selected from the segments. The envelope spectra are $\{E_{s1}, E_{s2}, \dots, E_{sp}\}$, and the corresponding JB series are $\{JB_{s1}, JB_{s2}, \dots, JB_{sp}\}$, then weighted narrowband envelope spectrum is calculated as

$$WNES(n) = \frac{\sum_{i=1}^p JB_{si} E_{si}(n)}{\sum_{i=1}^p JB_{si}} \quad (9)$$

3. EXPERIMENTAL INVESTIGATIONS

To verify the effectiveness of the proposed method in practical applications to enhanced bearing fault diagnosis, we established a simplified bearing test rig. The schematic diagram of the test rig is presented in figure 2. The test rig is made up of an induction motor, a shaft coupling to the motor, several bearings supporting the shaft, and an oil-loaded device as radial loader. The vibration signals at point 1, point 2 and point 3 are measured to simulate the situations with serious interference. The sampling frequency of the data acquisition system is 65536 Hz.

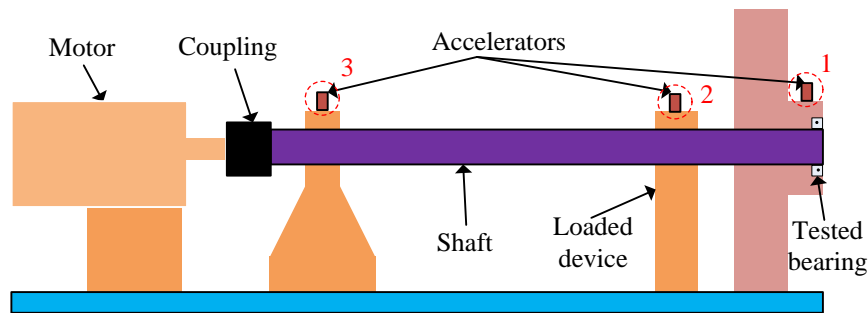


Figure 2. The schematic diagram of the test rig.

The type of tested bearings is 6010, and the parameters of the bearings are listed in table 1. Single point defect was seeded on the bearing race using wire cut electro-discharge machining (WEDM) to simulate inner-race and out-race defect. The outer-race fault characteristic frequency f_o and the inner-race fault characteristic frequency f_i are also presented in the table [15].

Table 1: Parameters of bearings (f_r is the shaft speed)

| Type | Pitch diameter(mm) | Ball diameter(mm) | Number of balls | f_i (Hz) | f_o (Hz) |
|------|--------------------|-------------------|-----------------|------------|------------|
| 6010 | 65 | 9 | 13 | $7.4f_r$ | $5.6f_r$ |

3.1. Inner race fault detection

The rotation speed is kept constant at 1500 rpm in the experiment, so the characteristic frequency of inner-race defect is 185 Hz. The traditional envelope analysis (EA) and the fast Kurtogram are introduced for comparison. The band with the largest SK is selected in the fast Kurtogram. The processing results of signal measured at point 1 are presented in figure 3. Since the accelerator is mounted near the tested bearing, the signal-to-noise ratio is high in this situation. It can be found that all the three methods can detect the characteristic frequency 185 Hz, which means all the methods are succeed in detecting the inner-race fault. EA and WNES have better performance than the fast Kurtogram.

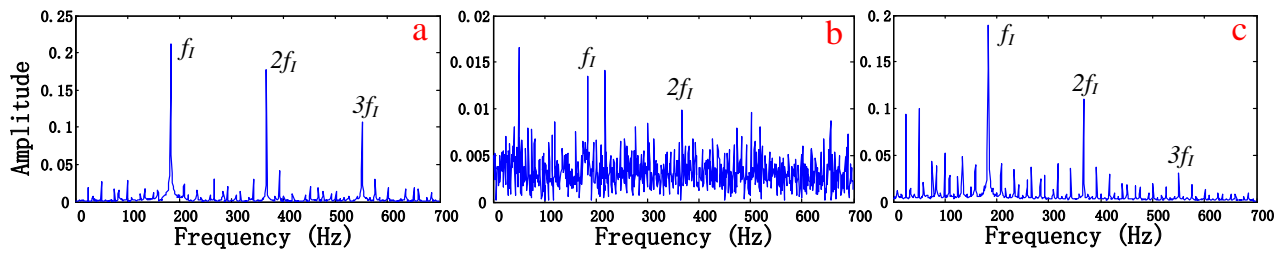


Figure 3. Analyzed results of signal with inner-race defect by different methods: (a) traditional EA; (b) fast Kurtogram; (c) WNES.

Signals are measured at point 2 and point 3 to simulate the situation with big interference, which is a universal phenomenon because sensors may not be mounted near the bearings in industrial fields. The processing results are presented in figure 4. Figure 4 (a)-(c) are spectra of signal measured at point 2, and (d)-(f) are results at point 3. The spectra obtained by EA are presented in figure 4(a) and figure 4(d). The spectra are contaminated by heavy noise and the characteristic frequency and its harmonics even did not appear in figure 4(d). The envelope waveforms demodulated by the fast Kurtogram are provided in figure 4(b) and (e). Both spectra show the presence of the fault signatures but not obviously. The weighted narrowband envelope spectra are presented in figure 4(c) and (f). It can be seen that the fault frequency and its harmonics are greatly enhanced and the spectra are pure. WNES is able to figure out characteristic frequency components in low signal-to-noise situation. The WNES has a better performance than EA and the fast Kurtogram in bearing inner-race fault detection.

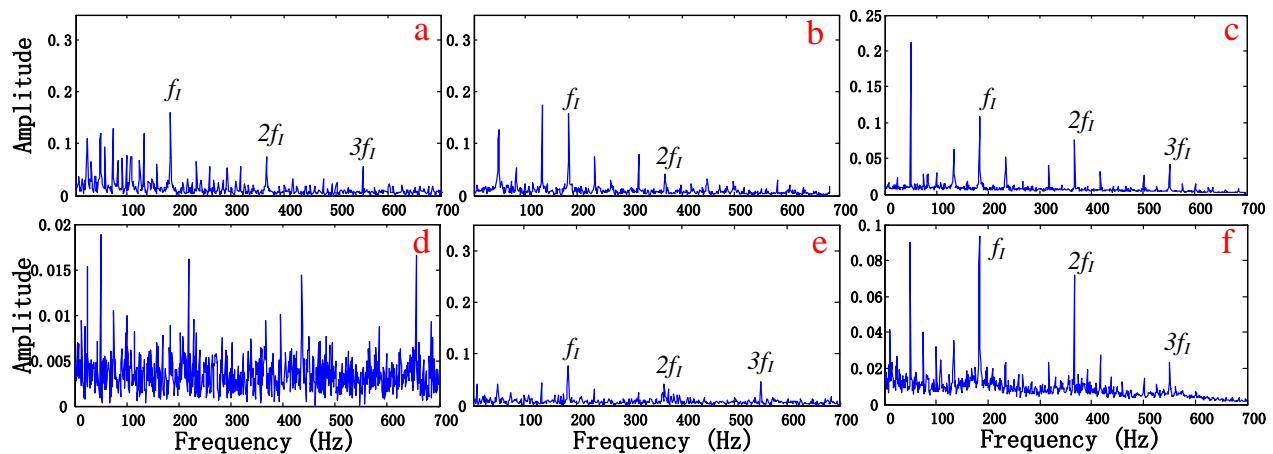


Figure 4. Analyzed results of signals with inner-race defect by different methods (a)-(c) Envelope spectra of signal measured at Point 2 demodulated by EA, fast Kurtogram and WNES (d)-(f) Envelope spectra of signal measured at Point 3 demodulated by EA, fast Kurtogram and WNES

3.2. Outer race fault detection

For this test, the rotating speed of the shaft is also fixed at 1500 rpm and the outer-race fault characteristic frequency is about 140 Hz. For the vibration signal measured at point 1, the spectra demodulated by traditional EA, the fast Kurtogram and WNES are presented in figure 5. The fault-related components are easily detected in the spectra. The three methods almost have the same performance. The fault is easy to be detected as the sensor attached near the tested bearing and the signal has a high signal-to-noise ratio.

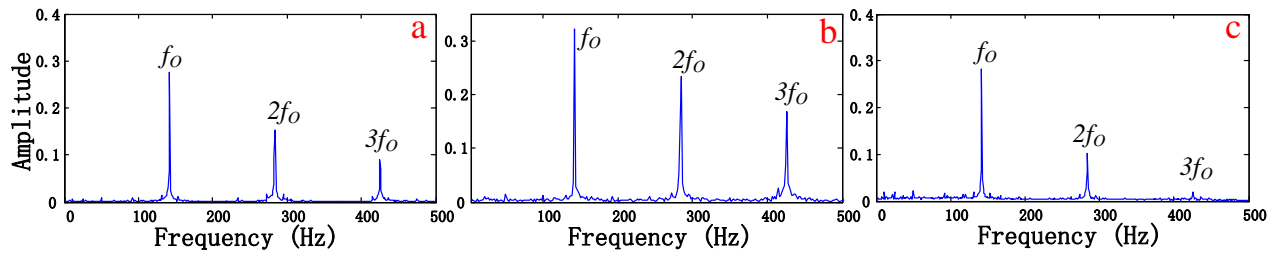


Figure 5. Analyzed results of signal with outer-race defect by different methods: (a) traditional EA; (b) fast Kurtogram; (c) WNES

For vibration signals measured at point 2 and point 3, the results are presented in figure 6. The spectra extracted by traditional EA are shown in figure 6(a) and (d), respectively. The envelope waveforms demodulated by the fast Kurtogram are provided in figure 6(b) and (e). Both the traditional EA and the fast Kurtogram are failed to detect the bearing characteristic frequency. The weighted narrowband envelope spectrum for signal measured at point 2 is presented in figure 6(c). Some other determined frequency components may affect the analysis, but the fault characteristic frequency f_o can be clearly identified. The weighted spectrum for signal measured at point 3 is presented in figure 6(f). Although the amplitudes of the characteristic frequency are small, the fault can also be successfully detected. These results indicate that the weighted narrowband envelope spectrum has an enhanced performance in bearing outer-race fault detection.

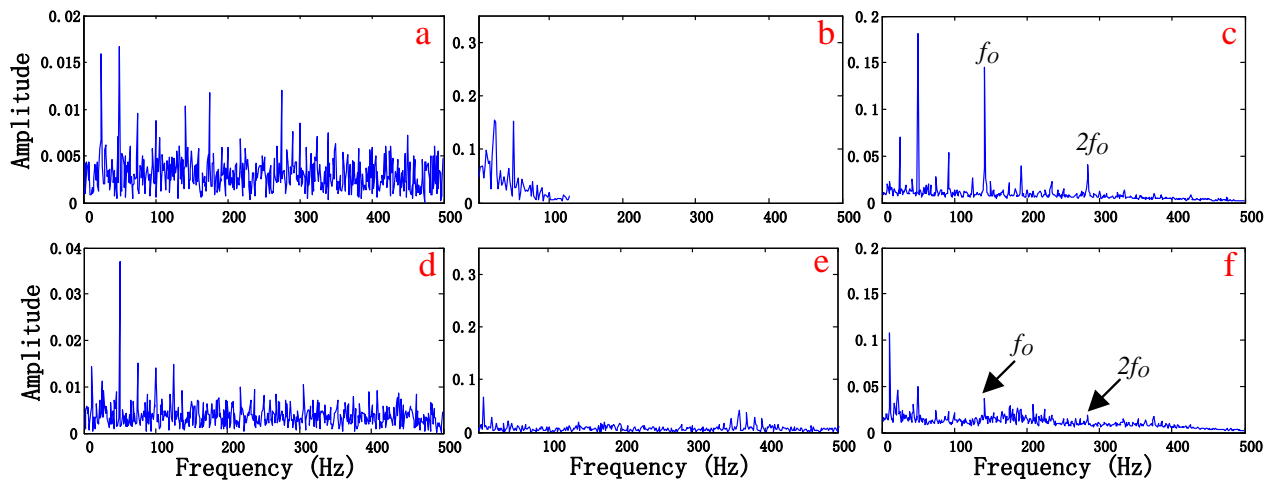


Figure 6. Analyzed results of signals with outer-race defect by different methods (a)-(c) Envelope spectra of signal measured at Point 2 demodulated by EA, fast Kurtogram and WNES (d)-(f) Envelope spectra of signal measured at Point 3 demodulated by EA, fast Kurtogram and WNES

The WNES has an ability to extract the bearing fault signatures in low signal-to-noise situation. As a result, after the comprehensive comparisons have been done, the proposed WNES shows the best performance for characteristic frequency identification in the cases studied. Therefore, the experimental results verified the effectiveness of the proposed method in bearing fault diagnosis.

4. CONCLUSIONS

In this study, the weighted narrowband envelope spectrum is proposed for enhanced bearing fault diagnosis. The method concerns the envelope spectra at multiple narrow bands, and the Jarque–Bera statistic of envelope spectrum amplitudes is introduced as index for the bands selection. WNES is obtained by weighted averaging of a sequence of envelope spectra at selected bands. The weighted spectrum does not wipe off all interference components and noises in the spectrum, but enhanced these determined components in the narrowband spectra. The proposed method has the ability for bearing fault diagnosis as the fault signatures are determined components in the envelope spectrum. Practical vibration signals acquired from bearings with an inner race defect and an outer race defect are used to validate the effectiveness and superiority of the proposed method. The results indicate that WNES has better

performance than traditional EA and the fast Kurtogram. Since determined components are preserved in the weighted envelope spectrum, it also has the potential to be used in situations containing two or more types of faults, which need to be further studied.

ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Foundation of China (approved Grant: 51575202, 51675204).

REFERENCES

- [1] Barszcz, T., & Jabłoński, A. (2011). A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mechanical Systems & Signal Processing*, 25(1), 431-451.
- [2] Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96, 289-306.
- [3] Li, B., Zhang, P. L., Wang, Z. J., Mi, S. S., & Liu, D. S. (2011). A weighted multi-scale morphological gradient filter for rolling element bearing fault detection. *Isa Transactions*, 50(4), 599-608.
- [4] Dwyer, R. (1983). Detection of non-Gaussian signals by frequency domain Kurtosis estimation. Paper presented at the Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP.
- [5] Antoni, J. (2006). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems & Signal Processing*, 20(2), 282-307.
- [6] Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems & Signal Processing*, 20(2), 308-331.
- [7] Antoni, J. (2007). Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems & Signal Processing*, 21(1), 108-124.
- [8] Sawalhi, N., Randall, R. B., & Endo, H. (2007). The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems & Signal Processing*, 21(6), 2616-2633.
- [9] Bozchalooi, I. S., & Liang, M. (2008). A joint resonance frequency estimation and in-band noise reduction method for enhancing the detectability of bearing fault signals. *Mechanical Systems & Signal Processing*, 22(4), 915-933.
- [10] Lei, Y., Lin, J., He, Z., & Zi, Y. (2011). Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems & Signal Processing*, 25(5), 1738-1749.
- [11] Wang, D., Tse, P. W., & Tsui, K. L. (2013). An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems & Signal Processing*, 35(1-2), 176-199.
- [12] Wang, Y., Xiang, J., Markert, R., & Liang, M. (2016). Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. *Mechanical Systems & Signal Processing*, s 66-67, 679-698.
- [13] Brown, A. O. S., & Jennifer. (2014). Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems & Signal Processing*, 48(1-2), 138-152.
- [14] Cong, F., Chen, J., Dong, G., & Zhao, F. (2013). Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis. *Mechanical Systems & Signal Processing*, 34(s 1-2), 218-230.
- [15] Cong, F., Zhong, W., Tong, S., Tang, N., & Chen, J. (2015). Research of singular value decomposition based on slip matrix for rolling bearing fault diagnosis. *Journal of Sound & Vibration*, 344, 447-463.

AUTHOR INDEX

| | | | |
|------------------------|---------------|-----------------------|-------------------------------------|
| <i>Abdulwahab, A</i> | 8 | <i>Dixon, D</i> | 163 |
| <i>Abufroukh, A</i> | 403 | <i>Dong, G</i> | 488, 505 |
| <i>Ahmed, Z J</i> | 270 | <i>Du, M</i> | 179, 192 |
| <i>Albarzenji, D</i> | 321 | <i>Duan, F</i> | 118 |
| <i>Alkali, B</i> | 368 | <i>Duan, J</i> | 534 |
| <i>Al-Khafaji, A</i> | 329 | <i>Eichentopf, B</i> | 278 |
| <i>Alnano, M</i> | 479 | <i>Eshaafi, T</i> | 355 |
| <i>Alseiari, A</i> | 517 | <i>Etebu, E</i> | 17 |
| <i>Anas, E R</i> | 497 | <i>Farrell, P</i> | 517 |
| <i>Arnaiz, A</i> | 98 | <i>Ferreiro, S</i> | 98 |
| <i>Asim, T</i> | 355 | <i>Fietz, N</i> | 278 |
| <i>Baglee, D</i> | 155, 163, 172 | <i>Fletcher, S</i> | 227 |
| <i>Ball, A D</i> | 41, 200, 309 | <i>Gao, L</i> | 179 |
| | 389 | <i>Gebler, O</i> | 256 |
| <i>Baltes, R</i> | 278 | <i>George, J</i> | 106 |
| <i>Bannett, I</i> | 118 | <i>Gonzalez, E</i> | 83 |
| <i>Barker, M</i> | 256 | <i>Gonzalez, M</i> | 287 |
| <i>Barros, , A</i> | 3 | <i>Gorostegui, U</i> | 155, 479 |
| <i>Berry, C</i> | 220 | <i>Gorton, S</i> | 227 |
| <i>Bravo, I</i> | 98 | <i>Grosvenor, R I</i> | 270, 329 |
| <i>Brethee, K</i> | 41, 309 | <i>Gu, F</i> | 41, 200, 237, 309, 337, 344, 389 |
| <i>Brooks, H</i> | 459 | | |
| <i>Campos, J</i> | 155 | <i>Guo, F</i> | 192 |
| <i>Cao, Z</i> | 192 | <i>Haba, U</i> | 309 |
| <i>Cassels, B</i> | 68 | <i>Hall, G</i> | 220 |
| <i>Chen, J</i> | 488, 505 | <i>Harrison, A</i> | 256 |
| <i>Chen, N</i> | 135 | <i>Hassani, F</i> | 303 |
| <i>Chen, P</i> | 396 | <i>Hassin, O</i> | 309 |
| <i>Chen, S</i> | 144 | <i>He, S</i> | 419 |
| <i>Chen, W</i> | 25 | <i>Hegedus, C</i> | 479 |
| <i>Cheng, Z</i> | 237 | <i>Hemmer, M</i> | 450 |
| <i>Clough, D</i> | 227 | <i>Hicks, B</i> | 256 |
| <i>Croes, J</i> | 287 | <i>Hidaka, R</i> | 428 |
| <i>Davies, B</i> | 526 | <i>Hu, J</i> | 144 |
| <i>de Calle, M</i> | 98 | <i>Hu, N</i> | 237 |
| <i>Deng, L</i> | 144 | <i>Huber, L G</i> | 377 |
| <i>Desmet, W</i> | 287 | <i>Hugo, A</i> | 211 |
| <i>Dettling, M</i> | 377 | <i>Ihemegbulem, I</i> | 172 |
| <i>Di Orio, G</i> | 479 | <i>Ikeda, M</i> | 92 |
| <i>Dinmohammadi, F</i> | 368 | <i>Ilyes, K</i> | 211 |

AUTHOR INDEX (Continued)

| | | | |
|--------------------------|--------------------|-------------------------|-------------------|
| <i>Irons, J</i> | 355 | <i>Prickett, P W</i> | 270 |
| <i>Jackson, L</i> | 526 | <i>Quentin, L</i> | 211 |
| <i>Jantunen, E</i> | 155, 479 | <i>Raadnui, S</i> | 361 |
| <i>Jiang, H</i> | 488 | <i>Rafezi, H</i> | 303 |
| <i>Kang, Y</i> | 337 | <i>Robbersmyr, K G</i> | 450 |
| <i>Kansanaho, J</i> | 185 | <i>Robson, K</i> | 163 |
| <i>Kärkkäinen, T</i> | 185 | <i>Ruan, J</i> | 135 |
| <i>Khan, M T I</i> | 428 | <i>Saarinen, K</i> | 185 |
| <i>Konde, E</i> | 98 | <i>Salgado, O</i> | 287 |
| <i>Kunz, S</i> | 377 | <i>Seyr, H</i> | 3 |
| <i>Lai, W</i> | 296 | <i>Shafiee, M</i> | 17, 368 |
| <i>Larrinaga, F</i> | 479 | <i>Shao, Y</i> | 179, 192 |
| <i>Leavers, V</i> | 76 | <i>Shark, L-K</i> | 68, 220 |
| <i>Li, H</i> | 344, 445 | <i>Sharma, P</i> | 155 |
| <i>Liang, X</i> | 118 | <i>Shen, Z</i> | 344 |
| <i>Liao, G</i> | 296 | <i>Sherrington, I</i> | 403, 459 |
| <i>Lin, J</i> | 52, 129 | <i>Shi, T</i> | 296, 534 |
| <i>Lin, L</i> | 433 | <i>Shi, Z</i> | 61, 337, 344, 445 |
| <i>Liu, F</i> | 200 | <i>Sinzara, W</i> | 459 |
| <i>Longstaff, A P</i> | 227 | <i>Smith, E H</i> | 459 |
| <i>Ma, J</i> | 61 | <i>Song, L</i> | 396 |
| <i>Malo, P</i> | 479 | <i>Song, Z</i> | 445 |
| <i>Mao, L</i> | 526 | <i>Swinton, J</i> | 355 |
| <i>Marquez, L</i> | 459 | <i>Tang, X</i> | 389 |
| <i>Mattonen-Arola, S</i> | 172 | <i>Tao, G</i> | 25 |
| <i>Matuszewski, B J</i> | 220, 497 | <i>Tautz-Weinert, J</i> | 247 |
| <i>Mba, D</i> | 118 | <i>Tsurumaru, S</i> | 428 |
| <i>Mein, S. J.</i> | 68 | <i>Turner, A</i> | 68 |
| <i>Melero, J</i> | 83 | <i>Usuda, T</i> | 92 |
| <i>Miao, Y</i> | 52 | <i>Van Vuuren, P A</i> | 389 |
| <i>Mishra, R</i> | 8, 321, 355 | <i>Varley, M</i> | 467 |
| <i>Mizushima, Y</i> | 396 | <i>Wang, D</i> | 433 |
| <i>Muskulus, M</i> | 3 | <i>Wang, Jianming</i> | 52, 129 |
| <i>Nienhaus, K</i> | 278 | <i>Wang, Jianguo</i> | 144 |
| <i>Nixon, A</i> | 68 | <i>Wang, L</i> | 179, 192 |
| <i>Onsy, A</i> | 106, 403, 459, 467 | <i>Wang, T I</i> | 450 |
| <i>Ota, S</i> | 396 | <i>Wang, Y</i> | 419 |
| <i>Philip, J</i> | 467 | <i>Wangveera, K</i> | 361 |
| <i>Pluymers, B</i> | 287 | <i>Watson, S J</i> | 247 |
| | | <i>Wei, Y</i> | 34 |

AUTHOR INDEX (Continued)

| | |
|-----------------------|--------------------|
| <i>Wheatley, A</i> | 163, 172 |
| <i>Xu, Xiaioqiang</i> | 129 |
| <i>Xu, Xiaoyu</i> | 61 |
| <i>Xu, Y</i> | 200, 389 |
| <i>Xuan, J</i> | 296, 534 |
| <i>Yan, C</i> | 129 |
| <i>Yanez, M J</i> | 459 |
| <i>Yang, Y</i> | 179, 192 |
| <i>Yin, L</i> | 179 |
| <i>Yu, R</i> | 135 |
| <i>Yuan, H</i> | 505 |
| <i>Yuan, J</i> | 34 |
| <i>Zhang, C</i> | 144 |
| <i>Zhang, H</i> | 61, 337, 344, 445 |
| <i>Zhang, L</i> | 237 |
| <i>Zhang, S</i> | 419 |
| <i>Zhang, X</i> | 135 |
| <i>Zhao, J</i> | 34 |
| <i>Zhao, M</i> | 52, 129 |
| <i>Zhen, D</i> | 61, 337, 344, 445, |
| <i>Zheng, C</i> | 179 |
| <i>Zhenqiao, Z</i> | 25 |
| <i>Zhipeng, H</i> | 25 |
| <i>Zhou, H</i> | 296, 534 |
| <i>Zhou, J</i> | 52, 129 |
| <i>Zhou, Yong</i> | 433 |
| <i>Zhou, Yu</i> | 34 |
| <i>Zurutuza, U</i> | 479 |